166

Journal of Educational Psychology

LIBRARY

Wayne H. Holtzman, Editor
University of Texas

Advisory Editors

Richard C. Anderson
University of Illinois

Daniel E. Berlyne
University of Toronto

C. Victor Bunderson
University of Texas

Lee J. Cronbach
Stanford University

Nathaniel Gage
Stanford University

Robert M. Gagné
Florida State University

Robert Glaser
University of Pittsburgh

Edmund W. Gordon
Teachers College, Columbia University

Ernest R. Hilgard
Stanford University

Philip W. Jackson
University of Chicago

Arthur R. Jensen
University of California, Berkeley

Herbert J. Klausmeier
University of Wisconsin

Nathan Kogan
Educational Testing Service, Princeton

C. Robert Pace
University of California, Los Angeles

David G. Ryans
University of Hawaii

Melvin Sikes
University of Texas

George G. Thompson
Ohio State University

Robert L. Thorndike
Teachers College, Columbia University

Joanna P. Williams
University of Pennsylvania

Bureau of Ednl. & Psyl. Research
(S. C. E. R. T.) 6.3.81
Acc 5/6 5675

VOLUME 62, 1971

Helen Orr
Managing Editor

Susan T. Hagerty
Technical Editor

Henry S. Salmon
Advertising and Business Manager

Anne Redman
Subscription Manager

PUBLISHED BIMONTHLY BY:
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth Street, N. W.
WASHINGTON. D. C. 20036

$10.50

Copyright, 1971, by the American Psychological Association, Inc.

CONTENTS OF VOLUME 62

Abramson, Theodore. See Tobias, Sigmund.

Ackerman, Margaret D. See Williams, Joanna P.

Ammon, Mary Sue. See Ammon, Paul R.

Ammon, Mary Sue. See Rohwer, William D., Jr.

Ammon, Paul R., and Ammon, Mary Sue. Effects of Training Black Preschool Children in Vocabulary versus Sentence Construction........ 421

Anderson, Richard C., Goldberg, Sheila R., and Hidde, Janet L. Meaningful Processing of Sentences........ 395

Anderson, Richard C., Kulhavy, Raymond W., and Andre, Thomas. Feedback Procedures in Programmed Instruction........ 148

Anderson, Richard C., Royer, James M., Kulhavy, Raymond W., Thornburg, Sylvia D., and Klemt, Laura L. Thematic Prompting in Paired-Associate Learning........ 315

Anderson, Richard C. See Monograph listing.

Anderson, Richard C. See Watts, Graeme.

Andre, Thomas. See Anderson, Richard C.

Andrews, O. Ernest. See Levinthal, Charles F.

Athanasiou, Robert. Selection and Socialization: A Study of Engineering Student Attrition........ 157

Barrett, Robert. See Rothbart, Myron.

Beane, William E., and Lemke, Elmer A. Group Variables Influencing the Transfer of Conceptual Behavior........ 215

Bergan, John R., Zimmerman, Barry J., and Ferg, Maureen. Effects of Variations in Content and Stimulus Grouping on Visual Sequential Memory........ 400

Berger, Stephen E. See Millham, Jim.

Berman, Phyllis, and Villwock, Mary Ann. Response Latency: Errors and Subsequent Change of Response on Retest........ 474

Beswick, David G., and Tallmadge, G. Kasten. Reexamination of Two Learning Style Studies in the Light of the Cognitive Process Theory of Curiosity........ 456

Blackhurst, A. Edward. See Tisdall, William J.

Blaine, Daniel D., and Dunham, Jack L. Effect of Availability on the Relationship of Memory Abilities to Performance in Multiple-Category Concept Tasks. 333

Blair, Claudia. See Gorfein, David S.

Blank, Marion, Higgins, Thomas, J., and Bridger, Wagner H. Stimulus Complexity and Intramodal Reaction Time in Retarded Readers........ 117

Bowers, Norman D. See Friedman, Philip.

Bridger, Wagner H. See Blank, Marion.

Brown, William F., Wehe, Nathan O., Zunker, Vernon G., and Haslam, Warren L. Effectiveness of Student-to-Student Counseling of the Academic Adjustment to Potential College Dropouts........ 285

Bruning, Robert H., and Lantinga, Larry J. Effects of Encounter with Verbal Units in Prose Contexts on Subsequent Associative Learning........ 308

Calfee, Robert C., and Jameson, Penny. Visual Search and Reading........ 501

Clark, D. Cecil. See Monograph listing.

Clemm, C. Kessler, iii. See Todd, Willliam B.

Coleman, Edgar. See Stevenson, Harold W.

Collins, James R. See Worthen, Blaine R.

Costin, Frank. Empirical Test of the "Teacher-Centered" versus "Student-Centered" Dichotomy........ 410

Cradler, John D., and Goodwin, Dwight L. Conditioning of Verbal Behavior as a Function of Age, Social Class, and Type of Reinforcement........ 279

Cromer, Ward. See Oaken, Robert.

Cromer, Ward, See Steiner, Rollin.

Dalfen, Susan. See Rothbart, Myron.

Deno, Stanley L., Jenkins, Joseph R., and Marsey, Judy. Transfer Variables and Sequence Effects in Subject-Matter Learning........ 365

Deno, Stanley L. See Jenkins, Joseph R.
Domino, George. Interactive Effects of Achievement Orientation and Teaching Style on Academic Achievement 427
Dunham, Jack L. See Blaine, Daniel D.
Epps, Edgar G., Katz, Irwin, Perry, Aubrey, and Runyon, Eugene. Effect of Race on Comparison Referent and Motives on Negro Cognitive Performance .. 201
Emmer, Edmund T., Good, Thomas L., and Oakland, Thomas D. Effect of Feedback Expectancy on Choice of Teaching Styles 451
Fazio, Anthony F., Lohr, Jeffrey M., and Screven, C. G. Reinforcement and Educationally Disadvantaged Boys: An Exploratory Study 245
Ferg, Maureen. See Bergan, John R.
Fischbach, Thomas J., and Walberg, Herbert J. Weighted and Unweighted Means for Estimation: A Note on the Humphreys-Dachler and Jensen Papers 79
Frase, Lawrence T. Effect of Incentive Variables and Type of Adjunct Question upon Text Learning 371
Friedman, Philip, and Bowers, Norman D. Student Imitation of a Rewarding Teacher's Verbal Style as a Function of Sex and Grade Level 487
Gallessich, June. See Schmidt, Linda.
Goldberg, Sheila R. See Anderson, Richard C.
Good, Thomas L. See Emmer, Edmund T.
Gordon, Leonard V. Weber in the Classroom 60
Gorfein, David S., and Blair, Claudia. Factors Affecting Multiple-Trial Free Recall 17
Goodwin, Dwight L. See Cradler, John D.
Grotelueschen, Arden. See McQuarrie, Duncan.
Guthrie, John T. Relationships of Teaching Method, Socioeconomic Status, and Intelligence in Concept Formation 345
Hall, William S. See Landis, Dan.
Hansen, Duncan N. See Leherissey, Barbara L.
Haslam, Warren L. See Brown, William F.
Hayman, John L., Jr. See Landis, Dan.
Hicks, Robert A., and Packwood, James. Effects of Structural Complexity of Bigrams on Paired-Associate Learning and Retention of Children 290
Hidde, Janet L. See Anderson, Richard C.
Higgins, Thomas J. See Blank, Marion.
Horvitz, James M. See Levin, Joel R.
Jacobson, Leonard I. See Millham, Jim.
Jameson, Penny. See Calfee, Robert C.
Jenkins, Joseph R., and Deno, Stanley L. Influence of Knowledge and Type of Objectives on Subject-Matter Learning 67
Jenkins, Joseph R. See Deno, Stanley L.
Jensen, Arthur R. Individual Differences in Visual and Auditory Memory 123
Jung, Steven M. Influence of Fading Procedures on the von Restorff Effect 445
Kaplan, Sandra A. See Levin, Joel R.
Katz, Irwin. See Epps, Edgar G.
Katz, Leonard, and Wicklund, David A. Word Scanning Rate for Good and Poor Readers 138
Khatena, Joe. Teaching Disadvantaged Preschool Children to Think Creatively with Pictures 384
Klemt, Laura L. See Anderson, Richard C.
Koran, Mary Lou. Differential Response to Inductive and Deductive Instructional Procedures 300
Koran, Mary Lou, Snow, Richard E., and McDonald, Frederick J. Teacher Aptitude and Observational Learning of a Teaching Skill 219
Kulhavy, Raymond W. See Anderson, Richard C.
Landis, Dan, Hayman, John L., Jr., and Hall, William S. Multidimensional Analysis Procedures for Measuring Self-Concept in Poverty Area Classrooms .. 95
Lansky, Leonard M. See Leventhal, Charles F.
Lantinga, Larry J. See Bruning, Robert H.

LEHERISSEY, BARBARA L., O'NEIL, HAROLD F., JR., AND HANSEN, DUNCAN N. Effects of Memory Support on State Anxiety and Performance in Computer-Assisted Learning.......... 413

LEMKE, ELMER A. See BEANE, WILLIAM E.

LEVIN, JOEL R., AND HORVITZ, JAMES M. The Meaning of Paired Associates.......... 209

LEVIN, JOEL R., HORVITZ, JAMES M., AND KAPLAN, SANDRA A. Verbal Facilitation of Paired-Associate Learning: A Limited Generalization.......... 439

LEVIN, JOEL R. See ROHWER, WILLIAM D., JR.

LEVINTHAL, CHARLES F., LANSKY, LEONARD M., AND ANDREWS, O. ERNEST. Student Evaluations of Teacher Behaviors as Estimations of Real-Ideal Discrepancies: A Critique of Teacher Rating Methods.......... 104

LIPE, DEWEY. See QUIRK, THOMAS J.

LOHR, JEFFREY M. See FAZIO, ANTHONY F.

LONG, K. KILBY. Transfer from Teaching to Learning.......... 167

LYNCH, STEVE, AND ROHWER, WILLIAM D., JR. Effects of Verbal and Pictorial Elaborations on Associative Learning and Response Learning in a Children's Paired-Associate Task.......... 339

MARKS, CLAUDE H. See TISDALL, WILLIAM J.

MARSEY, JUDY. See DENO, STANLEY L.

MCDONALD, FREDERICK J. See KORAN, MARY LOU.

MEANS, GLADYS H. See MEANS, ROBERT S.

MEANS, ROBERT S., AND MEANS, GLADYS H. Achievement as a Function of the Presence of Prior Information concerning Aptitude.......... 185

MCQUARRIE, DUNCAN, AND GROTELUESCHEN, ARDEN. Effects of Verbal Warning upon Misapplication of a Rule of Limited Applicability.......... 432

MILLER, MARTIN T. Instructor Attitudes toward, and Their Use of, Student Ratings of Teachers.......... 235

MILLHAM JIM, JACOBSON, LEONARD I., AND BERGER, STEPHEN E. Effects of Intelligence, Information Processing, and Mediation Conditions of Conceptual Learning.......... 293

MYROW, DAVID L. See Monograph listing.

OAKEN, ROBERT, WIENER, MORTON, AND CROMER, WARD. Identification, Organization, and Reading Comprehension for Good and Poor Readers.......... 71

OAKLAND, THOMAS D. See EMMER, EDMUND T.

O'LEARY, LAWRENCE R. Comparative Study of the Perceived Relevance of Material to be Learned and its Impact on the Performance of Culturally Deprived Junior College Students.......... 405

O'NEIL, HAROLD F., JR. See LEHERISSEY, BARBARA L.

PACKWOOD, JAMES. See HICKS, ROBERT A.

PERRY, AUBREY. See EPPS, EDGAR G.

QUIRK, THOMAS J., STEEN, MARGARET T., AND LIPE, DEWEY. Development of the Program for Learning in Accordance with Needs Teacher Observation Scale: A Teacher Observation Scale for Individualized Instruction.......... 188

RABINOVITCH, SAM. See WEINSTEIN, RHONA.

REILLY, DAVID H. Auditory-Visual Integration, Sex, and Reading Achievement.......... 482

RHODES, FEN. Team Teaching Compared with Traditional Instruction in Grades Kindergarten through Six.......... 110

ROHWER, WILLIAM D., JR., AND AMMON, MARY SUE. Elaboration Training and Paired-Associate Learning Efficiency in Children.......... 376

ROHWER, WILLIAM D., JR., AMMON, MARY SUE, SUZUKI, NANCY, AND LEVIN, JOEL R. Population Differences and Learning Proficiency.......... 1

ROHWER. WILLIAM D., JR. See LYNCH, STEVE.

ROSS, DOROTHEA. Retention and Transfer of Mediation Set in Paired-Associate Learning of Educable Retarded Children.......... 322

ROTHBART, MYRON, DALFEN, SUSAN, AND BARRETT, ROBERT. Effects of Teacher's Expectancy on Student-Teacher Interaction.......... 49

ROYER, JAMES M. See ANDERSON, RICHARD C.

RUNYON, EUGENE. See EPPS, EDGAR G.

RYCHLAK, JOSEPH F., AND TOBIN, THOMAS J. Order Effects on the Affective Learning Styles of Overachievers and Underachievers.......... 141

Scandura, Joseph M., and Voorhies, Donald J. Effect of Irrelvant Attributes and Irrelevant Operations on Rule Learning 352
Schmidt, Linda, and Gallessich, June. Adjustment of Anglo-American and Mexican-American Pupils in Self-Contained and Team-Teaching Classrooms 328
Screven, C. G. See Fazio, Anthony. F.
Shearer, James W. See Tallmadge, G. Kasten.
Snow, Richard E. See Koran, Mary Lou.
Spring, Carl. Perceptual Speed in Poor Readers 492
Steen Margaret T. See Quirk, Thomas J.
Steiner, Rollin, Wiener, Morton, and Cromer, Ward. Comprehension Training and Identification for Poor Readers and Good Readers 506
Stevenson, Harold W., Williams, Anne Marie, and Coleman, Edgar. Interrelations among Learning and Performance Tasks in Disadvantaged Children 179
Sticht, Thomas G. Failure to Increase Learning using the Time Saved by the Time Compression of Speech 55
Suzuki, Nancy. See Rohwer, William D., Jr.
Tallmadge, G. Kasten, and Shearer, James W. Interactive Relationships among Learner Characteristics, Types of Learning, Instructional Methods, and Subject Matter Variables 31
Tallmadge, G. Kasten. See Beswick, David G.
Tennyson, Robert D., and Woolley, F. Ross. Interaction of Anxiety with Performance on Two Levels of Task Difficulty 463
Thornburg, Sylvia D. See Anderson, Richard C.
Tisdall, Wiliam J., Blackhurst, A. Edward, and Marks, Claude H. Divergent Thinking in Blind Children 468
Tobias, Sigmund, and Abramson, Theodore. Interaction among Anxiety, Stress, Response Mode, and Familiarity of Subject Matter on Achievement from Programmed Instruction 357
Tobin, Thomas J. See Rychlak, Joseph F.
Todd, William B., and Clemm, C. Kesler, iii. Influence of Response Mode, Sex, Reading Ability, and Level of Difficulty on Four Measures of Recall of Meaningful Written Material 229
Torrence, E. Paul. Stimulation, Enjoyment, and Originality in Dyadic Creativity 45
Turner, Charles. Effects of Race of Tester and Need for Approval on Children's Learning 240
Villwock, Mary Ann. See Berman, Phyllis.
Voorhies, Donald J. See Scandura, Joseph M.
Walberg, Herbert J. See Fischbach, Thomas J.
Watts, Graeme, and Anderson, Richard C. Effects of Three Types of Inserted Questions on Learning from Prose 387
Wehe, Nathan O. See Brown, William F.
Weinstein, Rhona, and Rabinovitch, Sam. Sentence Structure and Retention in Good and Poor Readers 25
Wicklund, David A. See Katz, Leonard.
Wiener, Morton. See Oaken, Robert.
Wiener, Morton. See Steiner, Rollin.
Williams, Anne Marie. See Stevenson, Harold W.
Williams, Joanna P., and Ackerman, Margaret D. Simultaneous and Successive Discrimination of Similar Letters 132
Woolley, F. Ross. See Tennyson, Robert D.
Worthen, Blaine, R., and Collins, James R. Reanalysis of Data from Worthen's Study of Sequencing in Task Presentation 15
Zimmerman, Barry J. See Bergan, John R.
Zunker, Vernon G. See Brown, William F.

MONOGRAPHS

Anderson, Richard C., and Myrow, David L. Retroactive Inhibition of Meaningful Discourse 81
D. Cecil Clark. Teaching Concepts in the Classroom: A Set of Teaching Prescriptions Derived from Experimental Research 253

D 361
6..9..71
Lie
search.

JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 62 February 1971 Number 1

CONTENTS

Population Differences and Learning Proficiency.......... 1
WILLIAM D. ROHWER, JR., MARY SUE AMMON, NANCY SUZUKI, AND JOEL R. LEVIN

Reanalysis of Data from Worthen's Study of Sequencing in Task Presentation.. 15
BLAINE R. WORTHEN AND JAMES R. COLLINS

Factors Affecting Multiple-Trial Free Recall.......... 17
DAVID S. GORFEIN AND CLAUDIA BLAIR

Sentence Structure and Retention in Good and Poor Readers.......... 25
RHONA WEINSTEIN AND M. SAM RABINOVITCH

Interactive Relationships among Learner Characteristics, Types of Learning, Instructional Methods, and Subject Matter Variables.......... 31
G. KASTEN TALLMADGE AND JAMES W. SHEARER

Retroactive Interference in Reading Prose Materials.......... 39
JAMES H. CROUSE

Stimulation, Enjoyment, and Originality in Dyadic Creativity.......... 45
E. PAUL TORRANCE

Effects of Teacher's Expectancy on Student-Teacher Interaction.......... 49
MYRON ROTHBART, SUSAN DALFEN, AND ROBERT BARRETT

Failure to Increase Learning Using the Time Saved by the Time Compression of Speech.......... 55
THOMAS G. STICHT

Weber in the Classroom.......... 60
LEONARD V. GORDON

Influence of Knowledge and Type of Objectives on Subject-Matter Learning.... 67
JOSEPH R. JENKINS AND STANLEY L. DENO

Identification, Organization, and Reading Comprehension for Good and Poor Readers.......... 71
ROBERT OAKAN, MORTON WIENER, AND WARD CROMER

Weighted and Unweighted Means for Estimation: A Note on the Humphreys-Dachler and Jensen Papers.......... 79
THOMAS J. FISCHBACH AND HERBERT J. WALBERG

Monograph: *Retroactive Inhibition of Meaningful Discourse*.......... 81
RICHARD C. ANDERSON AND DAVID L. MYROW

© 1971 by the American Psychological Association, Inc.

Journal of Educational Psychology
1971, Vol 62, No. 1, 1-14

POPULATION DIFFERENCES AND LEARNING PROFICIENCY[1]

WILLIAM D. ROHWER, Jr.,[2] MARY SUE AMMON
University of California, Berkeley

NANCY SUZUKI AND JOEL R. LEVIN[3]
California State College, Long Beach
University of Wisconsin, Madison

The study was designed to evaluate the hypothesis that school-achievement deficiencies of low-socioeconomic-status black children are attributable to corresponding deficiencies in learning ability. Two populations (low-socioeconomic-status black and high-socioeconomic-status white) and three grade levels (kindergarten, first, and third) were sampled. The following tasks were administered: Peabody Picture Vocabulary Test, Children's Progressive Matrices, paired-associate learning, retention of paired associates, learning to learn in paired-associate tasks, and paired-associate learning under differing presentation methods. The results revealed large population differences on the Peabody Picture Vocabulary Test and the Children's Progressive Matrices at all grade levels; the largest difference was between the third-grade samples. In contrast, the paired-associate measures revealed a population difference only at the kindergarten level.

Currently, one of the most visible of educational phenomena is the marked discrepancy in school achievement between black children from families of low socioeconomic status and white children from high-socioeconomic-status families. It is commonly reported that differences between these two populations in performance on standardized achievement tests are as large as 40 to 50 points on percentile scales (Coleman, 1966; Rohwer, 1969; Wilson, 1963). The question is, What accounts for such observed differences in achievement?

One straightforward answer is that the two populations, high-socioeconomic-status white and low-socioeconomic-status black, differ in learning proficiency. If it is assumed that intelligence tests index learning proficiency in a relatively unbiased manner, then supporting evidence is readily and plentifully available. Differences in IQ between high-socioeconomic-status white and low-socioeconomic-status black children repeatedly have been shown to be in the same direction and of approximately the same magnitude as differences in standardized achievement test scores (Nichols, 1969). The problem with this explanation is the assumption that IQ indexes learning proficiency. Intelligence tests rarely require the child to engage in learning; they require him to give evidence that he has learned previously. Thus, rather than commanding immediate acceptance, the assumption needs empirical support.

The number of relevant studies presently available is very small. Few investigations have been undertaken in which learning tasks and intelligence tests have been administered to samples drawn from both high-socioeconomic-status white and low-socioeconomic-status black populations of school children. Those which have been conducted, however, cast considerable doubt on the validity of the assumption that IQ is an unbiased measure of learning proficiency. Semler and Iscoe (1963) observed the performance of white and black elementary school children on paired-associate learning tasks and on the Wechsler Intelligence Scale for Children (WISC). They found substantial race differences in WISC IQ but

---

[1] This work was supported, in part, by a contract with the United States Office of Economic Opportunity (B99-4776). The report was prepared at the Institute of Human Learning, which is supported by grants from the National Institutes of Health.

[2] Requests for reprints should be sent to William D. Rohwer, Jr., Institute of Human Learning, University of California, Berkeley, California 94720.

[3] The authors wish to acknowledge the indispensable assistance of the teachers and children in those schools where the data were collected as well as the contributions of Helgola Ross and Marsha Lynn in administering the various tests and of Douglas Penfield in analyzing the results.

not in paired-associate learning efficiency. Another paired-associate learning experiment (Rohwer, Lynch, Levin, & Suzuki, 1968) failed to detect differences between high-socioeconomic-status white and low-socioeconomic-status black children in any one of six variations of presentation method. Similarly, in a study confined entirely to black elementary school children, Green (1969) has reported finding no significant socioeconomic-status differences in paired-associate learning efficiency but marked socioeconomic-status differences in Lorge-Thorndike IQ. Jensen (1968) detected large differences between high-socioeconomic-status white and low-socioeconomic-status black children on the children's form of the Raven Progressive Matrices but not on a digit-span task. Direct measures of learning proficiency commonly fail to reveal population differences of precisely the kind that would be expected on the assumption that IQ is a valid and unbiased index of learning proficiency.

Nevertheless, the issue is not as simple as this brief review indicates. The relative performance of high-socioeconomic-status white and low-socioeconomic-status black children on intellectual tasks fluctuates as a function of a number of specific variables; chief among these are task differences and the chronological age of the subjects. When paired associates is the method used, the typical results are that significant amounts of between-groups variance are regularly associated with population membership among 3-, 4-, and 5-year-old children, occasionally in 6-year-olds, and rarely in children 7 years or older (Rohwer, 1967, Experiments XII and XIII; Rohwer et al., 1968; Rohwer & Lynch, 1968; Semler & Iscoe, 1963). The comparable developmental function has a very different form when digit-span tasks are used. In preschool children, that is, in 3-, 4-, and 5-year-olds, the performance of high-socioeconomic-status white and low-socioeconomic-status black children is virtually equivalent (Jensen, 1968), whereas in fourth-, fifth-, and sixth-grade children, digit memory among high-socioeconomic-status whites is markedly better than among low-socioeconomic-status blacks. Furthermore, digit-span performance is considerably better among high- than among low-socioeconomic-status black children at the third-grade level (Green, 1969).

In view of both the issue at stake and the evidence presently available, several questions deserve examination. Among these, the present study was addressed to the following ones.

Paired-Associate Test Reliability

The first is raised by the use of a learning task, in this case a paired-associate task, to estimate individual and group differences in learning efficiency. When a task is to be used for this purpose, it is important to know its reliability, but, in contrast to intelligence tests, such information is rarely available for learning tasks.

Population Differences and Varieties of Learning

Another question pertains to a hypothesis proposed by Jensen (1969) regarding a difference between high-socioeconomic-status and low-socioeconomic-status children in the organization of learning abilities. Jensen distinguishes two broad varieties of learning ability, associative (Level I) and conceptual (Level II). Presumably, Level I abilities are principally exercised on tasks that require the verbatim reproduction of the information originally presented for learning, tasks such as digit-span and paired-associate learning. In contrast, Level II abilities are elicited by tasks such as the Raven Progressive Matrices that require the subject to transform the information given in order to produce responses that are counted as being correct. On the assumption that Level I abilities are distributed equally among the two populations, high-socioeconomic-status white and low-socioeconomic-status black children, whereas Level II abilities are not, two interesting predictions follow. The first is that population differences should be detected on tasks that principally elicit Level II abilities but not on tasks that elicit Level I abilities (e.g., Raven versus paired-associate tests). The second prediction derived from the Jensen model is that the magnitude of the correlation between performance on Level I and

Level II tasks should be greater for high-socioeconomic-status white than for low-socioeconomic-status black children.

Learning to Learn

Do high-socioeconomic-status white and low-socioeconomic-status black children differ in the amount of transfer, namely, learning to learn, that accrues from performance on successive paired-associate lists? It might be argued that even if the two populations do not differ in single-list learning efficiency, they do differ in a capacity more vital for successful school learning, the capacity to transfer what has been learned from one instructional sequence to performance in another similar sequence.

Methods of Presentation

Are the results bearing on the issue of population differences specific to a particular method of presenting the pairs? In order to provide at least a limited answer to this question, the paired-associate test was constructed of lists of noun pairs within which the paired associates were presented in one or another of five different ways. Three of these item types were selected because of their demonstrable effect on learning efficiency. It has been shown that noun pairs depicted in the form of the objects to which they refer are learned more easily when they are (*a*) presented in the context of sentences, or (*b*) presented in the form of action episodes relating the two members of each pair, than when the objects are simply shown as still pictures and named aloud for the subject (Rohwer, Lynch, Levin & Suzuki, 1967).

Rohwer (1970b) has treated these methods of presentation as external analogues of hypothetical internal mental activities engaged in by persons who are efficient learners. The notion is advanced that successful paired-associate learning is promoted by the elaboration of the raw elements to be acquired so as to invest them with membership in a single semantic set, either by lodging them in the same linguistic unit, as in a sentence, or in the same pictorial unit, as in an action episode.

In addition to those forms of elaboration that can be construed as serving to form semantic sets, Rohwer (1968a) has also described two other forms that are more elementary in nature than the use of sentences or action imagery, but which are parallel in that one is verbal in character and the other is pictorial. The first of these primitive forms is that of generating a verbal label or name for pictorial stimuli and the second is that of generating a pictorial image of the referents of auditory stimuli.

The reason for manipulating the variable of paired-associate item types is that it permits a specification of the conditions under which population differences in learning efficiency occur, if they occur at all. Furthermore, it has been hypothesized (Rohwer, 1968a) that if low-socioeconomic-status black children have any deficiency in learning skills, it is a relatively weak propensity to elaborate the materials to be learned. From this hypothesis, the prediction follows that population differences are more likely to be detected on the less elaborated item types and not on the ones where the elaboration is furnished in the learning materials themselves.

Retention

The final question is concerned with the possibility of population differences in the efficiency with which information already learned can be recalled after a lapse of some specified amount of time. It might be reasoned that the inferior performance of low-socioeconomic-status black children on school achievement tests is due to limited capacity for initial learning, to limited capacity for retention, or to both. Thus provision was made for assessing the number of paired associates retained as well as the number initially learned.

METHOD

Subjects

The total sample numbered 288 children drawn in equal numbers from the six populations defined by the classification factors of grades (K, 1, 3) and populations (high-socioeconomic-status white, low-socioeconomic-status black). The populations were defined by the manner in which they were located as follows. The study was conducted in two communities known for the ethnic homogeneity of their school populations, one white, the other black.

TABLE 1

CHRONOLOGICAL-AGE MEANS AND STANDARD DEVIATIONS (IN MONTHS) AS A FUNCTION OF GRADES AND POPULATIONS

Grades	Populations			
	High-socioeconomic-status white		Low-socioeconomic-status black	
	$\bar{X}$	*SD*	$\bar{X}$	*SD*
Kindergarten	68.02	5.01	70.42	4.32
First	77.94	3.76	81.58	4.69
Third	102.62	4.32	106.25	4.74

Within the community from which the black sample was to be selected, a particular school was chosen in accord with the rule that it served a set of census tracts in which the households were clearly classifiable in terms of socioeconomic status. The variables available in census information were: median income, median education level, percentage homeowners, average value of homeowners' dwellings, average rent of other dwellings, ratio of "deteriorating" and dilapidated houses to "sound" houses, and a crowding index. After the schools were designated, sampling within grade levels was conducted by randomly selecting 24 males and 24 females from a list of all children enrolled in that grade. Within the groups originally selected, children absent on scheduled testing days were replaced from a list of randomly chosen alternates. There was a total of 7 such cases among the high-socioeconomic-status white samples and 19 among the low-socioeconomic-status black samples. Chronological age information for each of the samples is given in Table 1.

Tasks and Materials

Three different kinds of tasks were administered to every child individually: Peabody Picture Vocabulary Test, Form B; Coloured Raven Progressive Matrices; and, four paired-associated lists.

Peabody Picture Vocabulary Test. The test consists of a booklet with an array of four pictures appearing on every page. The subject's task is to point to the picture depicting the referent word uttered by the experimenter. The procedures as described in the manual (Dunn, 1965) were followed for both the administration and the scoring of the test. Thus, three measures were obtained for each child: raw score, mental age, and IQ. Based on the performance of the standardization sample, the alternate forms reliability of the test for children in Grades K, 1, and 3, respectively, are: .73, .69, .79.

Coloured Raven Progressive Matrices. The book form of the Coloured Raven Progressive Matrices was used, and, with a few exceptions, the procedures described in the manual (Raven, 1960) were followed in administering the test: the wording of the instructions was modified slightly to conform to American usage; and, prior to introducing the child to the first test problem in the book, he was given four practice problems in a board format. Although the manual does not provide information sufficient to lead to a confident estimate of the reliability of the test on populations such as those sampled here, Raven (1960, p. 15) notes that the test–retest coefficient is approximately .65 for children of the age range recruited for the present study.

Paired-associate test. Each of the four paired-associate lists is comprised of 25 noun pairs administered in accord with a study-test method for a total of two complete trials, that is, two study and two test trials. Within every list, the study-trial materials consist of the various pairs presented in one or another of five different ways, so that each of the five item types are represented by five pairs in each list. Since the types of items are distinguished from one another in terms of both auditory and visual features, they are presented by means of videotape to hold the test constant across administrations. The five item types are: nouns, in which each noun pair is presented aurally; pictures, in which each noun pair is represented by a picture of two objects; nouns-pictures, where pictures of object pairs are presented with the aural presentation of their labels; sentences-pictures, consisting of pictures of objects whose names are presented in a sentence describing some kind of interaction between them; and, nouns-action, where the visual signal literally depicts an interaction between the two objects shown while the names of the objects are presented aurally.

In order to permit an unequivocal attribution of expected differences among item types to corresponding differences among the presentation methods, a pretest was conducted to estimate the difficulty of each of the 100 pairs. Pairs were then assigned to lists and item types in a manner designed to equate for average difficulty.

The order of the pairs on the tape is random with respect to item type with the restriction that all types are represented once in each sequence of five pairs. During the study trials, successive pairs occur at a 4-second rate. The test trial materials for each list are also recorded on videotape. For every noun pair, either an object or a noun or both are presented. These stimuli appear at a 4-second rate, but in an order different from that of the study trials. As in the case of the study-trial materials, each item type is represented by a test stimulus in every sequence of five stimuli.

The instructions for the paired-associate task informed subjects about the various item types and urged them to learn each pair in such a way that they could supply the missing pair member on test trials. To clarify the instructions, a 5-item practice list, with one pair representing each of the five item types, was presented prior to each of the sets of two 25-item lists. The practice list was administered repeatedly until the subject attained a criterion of at least three correct responses.

Procedure

All subjects received the Coloured Raven Progressive Matrices, the Peabody Picture Vocabulary Test, and the paired-associate test during three separate testing sessions. The first session was devoted to the Coloured Raven Progressive Matrices, the second to the Peabody Picture Vocabulary Test and two of the paired-associate lists, and the third to the remaining two paired-associate lists. The first two sessions were separated by an interval of varying length, from 2 to 5 days, but in every case Sessions 2 and 3 were separated by a 2-day interval. The constancy of this latter interval is important because the third session always included the administration of the test trial materials from each of the two paired-associate lists learned in the second session. The purpose of this procedure was to assess paired-associate retention as a function of the various classification variables and of item types. Following the administration of the two new paired-associate lists, the third session concluded with the presentation of one test trial for each of the two paired-associate lists learned during the previous session.

Design

The analysis of variance design common to all three tasks was a three-way factorial, grades (K, 1, 3), populations (high-socioeconomic-status white, low-socioeconomic-status black), and sex (males, females). In the case of the paired-associate tasks, this basic design was augmented to permit the assessment of a number of sources of within-subjects variance. In designating these sources, it is necessary to distinguish between the dependent variables of original learning and recall. With respect to original learning, the additional variables were: item types (nouns, pictures, nouns-pictures, sentences-pictures, nouns-action); trials (1, 2); and, practice (first, second, third, and fourth lists). The variable of practice allows for an assessment of amounts of generalized transfer as a function of the subject classification variables of populations and grade level. It was possible to assess the effects of practice free of the influence of differences in difficulty among the four lists because the order in which the lists were administered was completely counterbalanced within each of the six samples. That is to say, two subjects in each sample were randomly assigned to each of the 24 possible list orders.

In addition to the status variables of grades, populations, and sex, the design for the analysis of the recall data included item types and lists (1, 2). The recall trials for each of the two lists were always administered in the order that the lists were presented during original learning. Once again, because of counterbalancing with respect to list order in original learning, all lists were equally represented in the first and second recall positions. It should be noted that neither the recall of the first nor of the second list can be construed as providing a measure of simple retention. Since the learning of the second, third, and fourth lists intervened between first-list learning and first-list recall, retention of the first list was subject to retroactive interference effects. Similarly, since the learning of the first-list preceded the learning of the second, second-list recall was subject to proactive interference effects as well as to possible retroactive effects from the learning of the third and fourth lists. Furthermore, second-list recall was also subject to possible interfering effects from the activity involved in the immediatly preceding attempt to recall the first list. Although the variety of possible interfering and facilitating effects complicates any interpretation of relative retention for first and second lists, between-samples contrasts for both measures are meaningful in view of the fact that all subjects were subject to the same effects.

The designs for the correlational analyses were straightforward. All, including both the reliability studies of the paired-associate test and the intertask correlations, were performed within the samples yielded by the combination of the factors, grades, and populations. The variables entered into these analyses were: Peabody Picture Vocabulary Test raw score, Coloured Raven Progressive Matrices raw score, paired-associate total score, and total scores for each of the five paired-associate item types.

RESULTS

Paired-Associate Test Reliability

The first aspect of the results to be examined is concerned with the reliability of the paired-associate test. The method of alternate forms was used to produce the reliability coefficients. For each of the six samples, six such coefficients were calculated, one for each of the item types and one for performance summed across item types. In every case, the scores consisted of the numbers of correct responses given on the test trials summed across two of the four lists. For all subjects, one form of the test was defined as the first two lists administered and the other form consisted of the remaining two lists. By this procedure, list differences were balanced across subjects. Thus, the maximum total score on either form of the test was 100 and the maximum score for each item type was 20.

The results are presented in Table 2 as a function of grades and populations. The reliability coefficients for the total score on the paired-associate test are acceptably high for most of the samples. In all cases, it may be that the coefficients reported underestimate the maximum reliability of some

TABLE 2

RELIABILITY COEFFICIENTS FOR THE PAIRED-ASSOCIATE TEST AS A FUNCTION OF GRADES, POPULATIONS, AND ITEM TYPES

Samples	Nouns	Pictures	Nouns-pictures	Sentences-pictures	Nouns-action	Total
High-SES white kindergarten	.45	.24	.28	.28	.32	.54
Low-SES black kindergarten	.59	.56	.57	.67	.76	.87
High-SES white first grade	.45	.56	.54	.42	.63	.80
Low-SES black first grade	.38	.52	.27	.31	.46	.67
High-SES white third grade	.50	.51	.50	.53	.44	.74
Low-SES black third grade	.62	.60	.58	.46	.47	.77

Note.—Abbreviated: SES = socioeconomic status.

particular pair of alternate forms available among the four paired-associate lists used. Even so, the reliability of the total score is quite comparable with the reliabilities reported for the Peabody Picture Vocabulary Test and the Coloured Raven Progressive Matrices.

As would be expected, the reliability coefficients produced by the individual item types are generally lower than those for total scores. When any one of the item types is treated as a test in itself, the factor of test length becomes important; each such test is only 10 items long, even when two of the full paired-associate lists are involved.

Population Differences and Varieties of Learning

Having established the relative comparability of the three tasks, paired associates, Peabody Picture Vocabulary Test, and Coloured Raven Progressive Matrices, with respect to reliability, the next matter of concern is to examine performance on the three as a function of grades, populations, and sex. Since the question of principal interest for each task was whether or not it detected a population difference at the three grade levels, the analyses of variance tested the simple main effects of populations within grades. For these analyses, the dependent variables for the Peabody Picture Vocabulary Test, the Coloured Raven Progressive Matrices, and the paired-associate tests, respectively, were correct items, correct solutions, and total correct responses. The means for these variables are presented in Table 3 and summaries of the three analyses of variance are given in Table 4. Because of the large numbers of hypotheses to be tested in the present study, the probability level for falsely rejecting each null hypothesis was set equal to .01.

An inspection of Table 3 suggests, and the information in Table 4 confirms, that for both populations and for all three tests, performance increases as a function of grade level. Note in Table 4 that the proportion of the total between-subjects sums of squares ($\hat{\omega}^2$) associated with grades seems approximately constant across the three tasks. In contrast, the value of $\hat{\omega}^2$ for populations appears to vary as a function of both grades and tests. In the cases of the Peabody Picture Vocabulary Test and the Coloured Raven Progressive Matrices, the proportion of the sums of squares attributable to population differences appears largest for the third-grade samples, whereas in the case of the paired-associate test that proportion seems to be smallest for the third-grade and largest for the kindergarten samples. Similarly, the total $\hat{\omega}^2$ for populations within grades, summed across grade levels, was larger for the Peabody Picture Vocabulary Test ($\hat{\omega}^2 = .32$) and the Coloured

TABLE 3

PERFORMANCE ON THE PEABODY PICTURE VOCABULARY TEST (PPVT), COLOURED RAVEN PROGRESSIVE MATRICES (CPM), AND PAIRED-ASSOCIATE TESTS AS A FUNCTION OF GRADES, POPULATIONS, AND SEX

Population	PPVT[a]			CPM[b]			Paired-associate test[c]		
	K	1	3	K	1	3	K	1	3
High-SES white									
Male	60.6	64.5	78.9	14.3	20.5	27.1	10.2	11.7	13.4
Female	58.5	63.2	76.1	15.5	19.5	26.6	8.4	10.0	13.2
Subtotal	59.6	63.8	77.5	14.9	20.0	26.8	9.3	10.8	13.2
Low-SES black									
Male	49.4	57.1	61.9	12.3	16.0	18.6	7.6	10.6	12.7
Female	46.0	52.2	59.2	13.2	13.8	16.1	6.7	9.4	11.4
Subtotal	47.7	54.6	60.5	12.8	14.9	17.3	7.2	10.0	12.0
Total	53.6	59.2	69.0	13.8	17.5	22.1	8.2	10.4	12.6

Note.—Abbreviated: K = Kindergarten, 1 = first grade, 3 = third grade, SES = socioeconomic status.

[a] Mean number of items correct.

[b] Mean number of correct problem solutions.

[c] Mean number of correct responses averaged across trials and lists (maximum possible score = 25).

TABLE 4

SUMMARIES OF ANALYSES OF VARIANCE PERFORMED ON RESULTS PRODUCED BY THE PEABODY PICTURE VOCABULARY TEST (PPVT), COLOURED RAVEN PROGRESSIVE MATRICES (CPM), AND PAIRED-ASSOCIATE TESTS

Source of variance	df	PPVT			CPM			Paired associate		
		MS	*F*	$\hat{\omega}^2$	*MS*	*F*	$\hat{\omega}^2$	*MS*	*F*	$\hat{\omega}^2$
Grades	2	5815.1	118.1*	.30	1641.8	110.3*	.31	744.7	66.9*	.30
Populations (A)										
Kindergarten	1	3384.4	68.7*	.09	106.3	7.1*	.01	176.0	15.8*	.04
First	1	2035.0	41.3*	.05	620.2	41.7*	.06	28.4	2.5	—
Third	1	6936.0	140.8*	.18	2175.5	146.2*	.21	58.0	5.2	—
Sex (B)										
Kindergarten	1	176.0	3.6	—	25.0	1.7	—	70.4	6.3	—
First	1	234.4	4.8	—	57.0	3.8	—	87.6	7.9*	.02
Third	1	181.5	3.7	—	55.5	3.7	—	25.0	2.2	—
A × B										
Kindergarten	1	9.4	—	—	.3	—	—	7.7	—	—
First	1	77.0	1.6	—	9.3	—	—	1.8	—	—
Third	1	.1	—	—	23.0	1.6	—	12.8	1.2	—
Error	276	49.2			14.9			11.1		

* $p < .01$.

Raven Progressive Matrices ($\hat{\omega}^2 = .28$) than for the paired-associate test ($\hat{\omega}^2 = .05$). Indeed, among the *F* ratios for populations, only that for the kindergarten samples was significant on the paired-associate test while *F*s were significant for all three grade levels on both the Peabody Picture Vocabulary Test and the Coloured Raven Progressive Matrices Test.

With one qualification, the results examined so far are consistent with the hypothesis advanced by Jensen (1969) concerning population differences in learning ability. The performance of high-socioeconomic-status whites substantially exceeded that of low-socioeconomic-status blacks on both of the tasks (Peabody Picture Vocabulary Test and Coloured Raven Progressive Matrices) that fall into Jensen's Level II category and the population difference was considerably less substantial on the task that is presumably of the Level I variety (paired associates). The qualification, of course, pertains to the fact that the magnitude of population differences seems to vary considerably with the ages of the subjects sampled. The Level II tasks appeared to yield more variance associated with populations at the third-grade than at the kindergarten level, and even the Level I task revealed a significant population difference for the kindergarten samples. Thus, within certain age limits, the present results confirm the first prediction derived from Jensen's hypothesis.

In contrast, the second prediction is clearly disconfirmed by the results of correlational analyses for the three tests administered in the present study. The prediction, it will be recalled, was that the magnitude of the correlation between Level I and Level II tasks should be larger among high-socioeconomic-status white than among low-socioeconomic-status black subjects. To assess this prediction, correlation coefficients between the paired-associate test and the Peabody Picture Vocabulary Test and Coloured Raven Progressive Matrices were calculated separately for each of the six samples. The results are presented in Table 5. An inspection of the values shown in Table 5 indicates that there is no support in these data for the notion that Level I and Level II abilities are more closely related in high-socioeconomic-status white than in low-socioeconomic-status black populations at any of the grade levels sampled.

TABLE 5

Product-Moment Correlation Coefficients between Scores on the Paired-Associate Test and Scores on the Peabody Picture Vocabulary Test (PPVT) and the Coloured Raven Progressive Matrices (CPM) as a Function of Grades and Populations

Grades	Populations			
	High-SES white		Low-SES black	
	PPVT	CPM	PPVT	CPM
Kindergarten	.47*	.12	.66*	.46*
First	.28	−.08	.35	.02
Third	.14	.00	.38*	.29

Note.—SES = socioeconomic status.
* $p < .01$.

Learning to Learn

For the various purposes attached to the assessment of within-subjects effects, scores on each of the four lists, five item types, and two trials were transformed into 39 new dependent variables and were subjected to multivariate analysis of variance in the manner suggested by Morrison (1967). The interactions of specified within-subjects variables and the between-subjects sources of variation were also examined. With respect to the issue of nonspecific transfer or learning to learn, the variates of interest are the scores obtained on each of the successive four paired-associate lists administered to every subject. The means for these variables are presented in Figure 1 as a function of grades and populations. As an examination of these data suggest, the analysis revealed a significant learning to learn effect ($F = 26.69$, $df = 3/274$, $p < .01$). A trend analysis confirmed the impression of improvement in performance across lists in that the linear component was significant (step-down $F = 39.66$, $df = 1/276$, $p < .01$). Although the quadratic component was not significant (step-down $F < 1$), the uniform drop in performance between the second and third lists administered was detected in the significant cubic component of the trend (step-down $F = 34.62$, $df = 1/276$, $p < .01$). In this connection it should be recalled that a 48-hour interval elapsed between the administration of the second and third lists. Thus, the drop in performance should probably be attributed to the loss of the benefit of warm-up across the interval. The total amount of improvement in performance from List 1 to List 4, that is, the total amount of nonspecific transfer observed, may be partitioned into two components—warm-up and learning to learn. The best estimate of the learning to learn component is the difference between Lists 2 and 4.

With respect to the question of principal interest for the learning to learn analysis, there was no significant interaction between populations and practice at any of the three grade levels (all Fs < 1). Nor were any of the interactions of sex with practice significant. Accordingly it is warranted to conclude that these data provide no support for the hypothesis that the high-socioeconomic-status white samples profit more than the low-socioeconomic-status black samples from previous experience with the kind of learning task administered. Indeed, the direction of the differences between second- and fourth-list performance appears to favor the low-socioeconomic-status black children (mean difference = .97 items), not the high-socioeconomic-status white children (mean difference = .10 items).

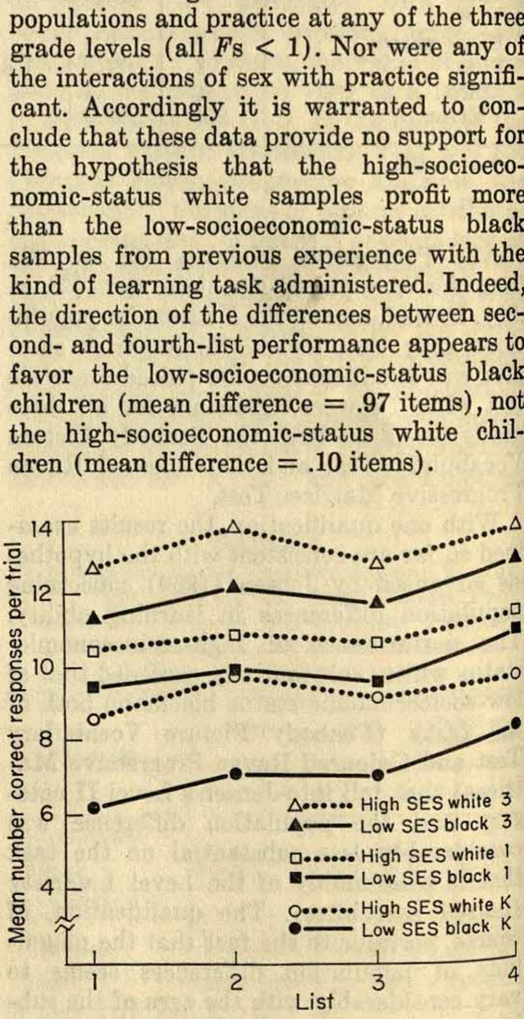


Fig. 1. Mean numbers of correct responses on the paired-associate task as a function of grades, populations, and practice (lists).

Methods of Presentation

Differences among the original item type variables are displayed in Figure 2 as a function of grades and populations. For the multivariate analysis designed to estimate the amount of variance attributable to these differences among methods of presenting the various paired associates, four orthogonal linear transformations of the five item type variates were made. The effect of item types was significant ($F = 1291.98$, $df = 4/273$, $p < .01$). Appropriate post hoc procedures revealed that all pairwise comparisons were significant. Another multivariate test of the transformed variables indicated that the effect of grades across the five variates was not constant ($F = 7.44$, $df = 8/546$, $p < .01$); the superiority of picture over noun items appears to increase with grades and the superiority of sentence-picture and noun-action items over noun, picture, and noun-picture items also appears to increase with grades. It is worth emphasizing the substantial magnitude of the item types effect since the estimates of learning efficiency produced by the present study vary so markedly with the method by which the paired associates were presented.

With respect to the question whether or not population differences in learning efficiency depend upon the manner in which paired associates are presented, the results indicate that the answer varies with the grade level sampled. In the kindergarten samples, the magnitude of the populations difference varies significantly across item types ($F = 3.52$, $df = 4/273$, $p < .01$); descriptively, the effect is that the superiority of the high-socioeconomic-status white sample is greater for sentence-picture and noun-action items than for noun, picture, and noun-picture items. Although the populations effect does not differ significantly across item types in the first-grade samples ($F = 2.99$, $df = 4/273$, $p = .019$), the direction of the differences appears to indicate that the population difference is smaller for the nouns-pictures items than for the nouns and for the pictures items. A similar pattern of results for populations across item types was detected for the third-grade samples ($F = 4.24$, $df = 4/273$, $p < .01$); population differences were larger for nouns and for pictures items than for nouns-pictures items. Indeed, an inspection of Figure 2 reveals that the mean differences on nouns-pictures items in both Grade 1 and in Grade 3 favor the low-socioeconomic-status black samples. In summary, it must be concluded that the detection of population differences in paired-associate learning efficiency varies significantly with the method of presentation employed.

Neither the effects associated with the factor of sex nor those associated with the interaction of sex and populations varied

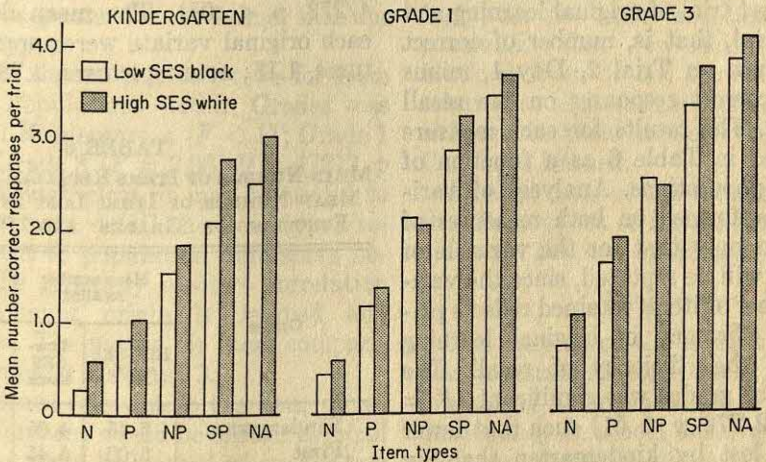


FIG. 2. Mean numbers of correct responses on the paired-associate task as a function of grades, populations, and item types.

significantly across the five item types variates. This result holds for all three of the grade levels sampled.

It will be recalled that each list was administered for a total of two trials. An analysis of the transformed variable, Trial 2 score minus Trial 1 score, revealed several interesting effects. This difference itself was significant ($F = 3454.85$, $df = 1/276$, $p < .01$), and the amount of the difference varied with grades ($F = 26.09$, $df = 2/276$, $p < .01$), such that the gain in correct responses from Trial 1 to Trial 2 increased with grade level. The magnitude of gain also varied significantly as a function of population within kindergarten ($F = 10.68$, $df = 1/276$, $p < .01$) and Grade 3 ($F = 8.79$, $df = 1/276$, $p < .01$) but not in Grade 1 ($F = 3.82$, $df = 1/276$, $p = .052$). In the case of each of the significant effects, the high-socioeconomic status white samples appear to gain more from trial to trial than the low-socioeconomic-status black samples. Trials did not differ significantly as a function of the factor of sex or as a function of the interaction of populations and sex at any of the three grade levels.

Retention

The efficiency of recall for the first two paired-associate lists administered after a 2-day retention interval was indexed in two ways: number of correct responses given on the recall trials; and amount lost between the second test trial of original learning and the recall trial, that is, number of correct responses given on Trial 2, Day 1, minus number of correct responses on the recall trial, Day 2. The results for each measure are presented in Table 6 as a function of grades and populations. Analyses of variance were performed on both measures of retention but only that for the variable of amount lost will be reported, since the variable of number of items retained reflects primarily the efficiency of original learning rather than the efficiency of recall. The main effect of grades was significant ($F = 31.06$, $df = 2/276$, $p < .01$) such that fewer items were lost by kindergarten than by first- and third-grade children; the means for the first- and third-grade samples did not differ significantly. The effect of populations was significant for the kindergarten samples ($F = 11.03$, $df = 1/276$, $p < .01$) but not for Grade 1 ($F = 2.03$, $df = 1/276$, $p > .05$) or for Grade 3 ($F < 1$). As can be seen in Table 6, there is evidence of more forgetting on the part of the high-socioeconomic-status white kindergarten children than on the part of the low-socioeconomic-status black children. Clearly, there is no evidence in these data to support the supposition that low-socioeconomic-status black children are deficient in their capacity for retaining what they learn.

The amount of variance in forgetting associated with sex was not significant for any of the three grade levels. The interaction of populations with sex, however, was significant for the kindergarten samples ($F = 8.63$, $df = 1/276$, $p < .01$) and not for the other two grades (both Fs < 1). Descriptively, in the kindergarten samples the form of the interaction is such that for high-socioeconomic-status whites, more items were lost by males than by females, whereas for low-socioeconomic-status blacks, more items were lost by females than by males.

The decreases in number of correct responses from Trial 2 of original learning to the recall trial for each of the item types were transformed into four new variables. The multivariate test for equality of these decreases was significant ($F = 115.29$, $df = 4/273$, $p < .01$). The mean decreases for each original variate were: nouns, .67; pictures, 1.13; nouns-pictures, 1.78; sentences-

TABLE 6

MEAN NUMBER OF ITEMS RECALLED ON DAY 2 AND MEAN NUMBER OF ITEMS LOST BY DAY 2 AS A FUNCTION OF GRADES AND POPULATIONS

Grades	Mean number recalled		Mean number lost	
	High-SES white	Low-SES black	High-SES white	Low-SES black
Kindergarten	5.45	4.05	6.25	4.62
First	6.00	5.45	7.61	6.92
Third	8.10	6.60	8.34	7.86

Note.—SES = socioeconomic status.

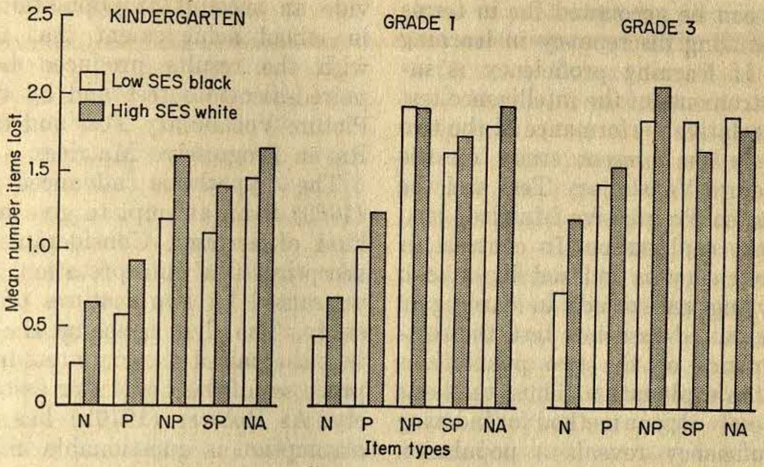


FIG. 3. Retention of paired associates (mean number of items lost) as a function of grades, populations, and item types.

pictures, 1.58; nouns-action, 1.77. Appropriate post hoc procedures revealed no significant differences among the nouns-pictures, sentences-pictures, and nouns-action item types (although each of these differed significantly from both nouns and pictures). This outcome is important in connection with the problem presented by the generally positive correlation between the variable of amount learned and that of amount lost. This pattern is broken in the present analysis where no more pairs were lost in the sentences-pictures than in the nouns-pictures item types, even though the number of correct responses during Trial 2 of original learning was greater for the former than for the latter types.

No one of the multivariate tests for Item Types × Populations within Grades was significant: Kindergarten ($F < 1$); Grade 1 ($F < 1$); Grade 3 ($F = 3.06$, $df = 4/273$, $p = .017$). Nevertheless, it may be useful to examine in more detail the item types results relevant to population differences because of the generally positive correlation between amount originally learned and amount lost. The means for these comparisons are depicted in Figure 3.

The evidence relevant to the issue of retention differences would be clearer if recall comparisons could be located where the two populations were equivalent at the end of original learning. An inspection of Figure 2 reveals two such instances: performance on nouns-pictures items in the first- and third-grade samples. (Although Figure 2 displays mean performance across trials, it accurately reflects Trial 2 performance as well). If low-socioeconomic-status black children are in fact deficient in their capacity to recall material previously learned, despite the overall results of the present study to the contrary, this deficiency should exhibit itself in higher mean loss scores for the first- and third-grade samples on the nouns-pictures item type. The data displayed in Figure 3, however, offer no support whatever for this hypothesis. Accordingly, for the task used here, it must be concluded that low-socioeconomic-status black children do not show a deficiency in relation to high-socioeconomic-status white children in their capacity to retain what they have learned.

Finally, none of the item types tests for the effects of sex or for the interaction Population × Sex was significant.

DISCUSSION

The purpose of the present investigation was to establish some facts necessary for evaluating an explanation of the observed discrepancy in school achievement between high-socioeconomic-status white and low-socioeconomic-status black children. At issue is the question whether or not this discrepancy can be accounted for in terms of a corresponding discrepancy in learning proficiency. If learning proficiency is indexed by instruments of the intelligence test variety, the relative performance of the two populations in the present study on the Peabody Picture Vocabulary Test and the Coloured Raven Progressive Matrices lend support to the explanation. In contrast, if learning proficiency is indexed by a task that directly engages subjects in learning, in this case the paired-associate test, the relative performance of the two populations contradicts the explanation. Thus, the issue remains unresolved; one method for indexing learning proficiency reveals a population discrepancy consistent with that observed for school achievement while another method finds no such discrepancy.

In examining some attempts to reconcile these disparate outcomes, it will be useful to recall the present results in detail. The data clearly indicate that the paired-associate test yields scores approximately equivalent in reliability to those yielded by the Peabody Picture Vocabulary Test and Coloured Raven Progressive Matrices, the results also confirm the inference from previous research that the magnitude of population differences depends on both task variables and chronological age. On the Peabody Picture Vocabulary Test and the Coloured Raven Progressive Matrics, the performance of high-socioeconomic-status white children exceeded that of low-socioeconomic-status black children at all grade levels sampled, kindergarten, first, and third, but the amount of variance associated with the population difference appeared larger for third-grade than for kindergarten children. On the paired-associate test, however, the population difference was significant only for the kindergarten samples. Again with respect to the paired-associate test, when interactions with item types are ignored, low-socioeconomic-status black children are *not* deficient relative to high-socioeconomic-status white children in either the efficiency of original learning or in the retention of what has been learned or in amount of nonspecific transfer, that is, in the degree to which they benefit from previous learning. Thus, the problem is to provide an account of population differences in school achievement that is consistent with the results produced both by the paired-associate test and by the Peabody Picture Vocabulary Test and the Coloured Raven Progressive Matrices.

The hypothesis advanced by Jensen (1969) is an attempt to give precisely this kind of account. Considerable caution in accepting this interpretation, however, is warranted by two features of the present study. The first concerns the assumption that the paired-associate test mainly elicits processes of the Level I or associative variety. As Rohwer (1970b) has argued, this assumption is questionable in view of the evidence that paired-associate learning involves considerable conceptual activity (Bower, 1968, 1969; Bugelski, 1962; Martin, 1967; Martin, Boersma, & Cox, 1965; Montague & Wearing, 1967; Paivio, 1967; Paivio, Yuille, & Smythe, 1966; Rohwer, 1967, 1968a, 1968b, 1970a; Rohwer & Levin, 1968; Rohwer & Lynch, 1966, 1967; Rohwer, Lynch, Levin, & Suzuki, 1967, 1968; Rohwer, Lynch, Suzuki, & Levin, 1967; Rohwer, Shuell, & Levin, 1967; Runquist & Farley, 1964). To be sure, paired-associate tasks require verbatim reproduction of responses, but this characterization of the performance demanded does not necessarily imply that paired associates are learned by rote processes. Another reason for caution in accepting the Level I–Level II account is raised by the results of the correlational analyses presented here. It will be recalled that from the model, Jensen (1969) derives the prediction that performance on Level I tasks will be more highly correlated with performance on Level II tasks among high- than among low-socioeconomic-status children. If the designations of the Peabody Picture Vocabulary Test and the Coloured Raven Progressive Matrices as Level II tasks and of the paired-associate test as a Level I test are accepted, then the correlational results presented here run directly counter to the prediction. Thus, on both counts the Level I–Level II model fails to provide a satisfactory explanation of population differences in school achievement.

Among alternative ways of accounting for this phenomenon, one obvious possibility is that paired-associate tasks do not elicit the kinds of learning processes necessary for successful performance on school learning tasks or on IQ tests. Clearly, this answer is subject to empirical evaluation; one method of assessing it is to compare the validity of the three tests administered in the present investigation for predicting performance on school achievement tests. Such a study is presently in progress. Meanwhile, caution should be exercised in concluding that paired-associate learning is irrelevant to school learning in view of the substantial relationships reported by Stevenson, Hale, Klein, and Miller (1968) between paired-associate learning and school achievement.

An alternative hypothesis has been proposed by Rohwer (1969). It begins by noting that any instrument which demands that the testee recall previous learnings will inevitably reveal population differences unless the degree of original learning has been equivalent among the populations assessed. If it is granted that high-socioeconomic-status white children achieve higher degrees of mastery than low-socioeconomic-status black children by the end of any given instructional unit, then any tests that probe for the recall of material learned from that unit will show a population difference. This category of instruments would include the Peabody Picture Vocabulary Test and virtually any standardized test of school achievement. This analysis, however, does not account for the fact that population differences have been observed on some tasks that principally require new learning as well as on tasks that require the recall of previous learnings. One example is provided by the results of the Coloured Raven Progressive Matrices presented here; other examples include performance on tasks as straightforward as digit span and the free recall of lists of familiar objects that are subsumable in formal categories. Rohwer (1969) has suggested that population differences on tasks such as these are attributable to a common property: efficient performance on each task requires the mastery of sets of formal conventions (for example, numbers and categories) created by cultural consensus that may be more readily available to or more valued by one population than by another. One implication of this position is that tasks where proficient performance depends more on skill at the application of imaginative, idiosyncratic conceptual processes than upon conventional formal processes will reveal equivalence of learning efficiency.

This hypothesis directs attention to the conditions of original learning, for it is these conditions, to a large extent, that determine the degree of learning that will be achieved and the manner of its achievement. In this connection, the results of the present study are provocative in three respects. First, they suggest the capacity of low-socioeconomic-status black children to recall previous learnings is as great as that of high-socioeconomic-status white children. Second, if it is assumed that successful performance on paired-associate tasks involves the operation of imaginative conceptual processes, the results are consistent with the expectation that population equivalence should be observed. Third, they make it clear that a truly remarkable amount of variance in the success of original learning, for both populations, is associated with item types, that is, with the manner in which the learning materials are presented.

REFERENCES

BOWER, G. H. Organizational factors in memory. Research Seminar, Institute of Human Learning, University of California, Berkeley, 1968.

BOWER, G. H. Mental imagery and associative learning. Paper presented at the Fifth Annual Symposium on Cognition, Carnegie-Mellon University, April 1969.

BUGELSKI, B. R. Presentation time, total time, and mediation in paired-associate learning. *Journal of Experimental Psychology*, 1962, **63**, 409–412.

COLEMAN, J. S., ET AL. *Equality of educational opportunity*. Washington, D.C.: United States Government Printing Office, 1966.

DUNN, L. M. *Expanded manual: Peabody Picture Vocabulary Test*. Minneapolis, Minn.: American Guidance Service, 1965.

GREEN, R. B. SES differences on learning and ability tests in black children. Unpublished master's thesis, University of California, Berkeley, 1969.

JENSEN, A. R. Patterns of mental ability and socioeconomic status. *Proceedings of the National Academy of Sciences*, 1968, **60**, 1330–1337.

JENSEN, A. R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, **39**, 1–123.

MARTIN, C. J. Associative learning strategies employed by deaf, blind, retarded and normal children. *Educational Research Series,* 1967, **38,** 1–158.

MARTIN, C. J., BOERSMA, F. J., & COX, D. L. A classification of associative strategies in paired-associate learning. *Psychonomic Science,* 1965, **3,** 455–456.

MONTAGUE, W. E., & WEARING, A. J. The complexity of natural language mediators and its relation to paired-associate learning. *Psychonomic Science,* 1967, **7,** 135–136.

MORRISON, D. F. *Multivariate statistical methods.* New York: McGraw-Hill, 1967.

NICHOLS, R. C. Implications of racial differences in intelligence for educational research and practice. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, February 1969.

PAIVIO, A. *Meaning, mediation, and memory.* (Research Bulletin No. 48) London, Canada: Department of Psychology, University of Western Ontario, 1967.

PAIVIO, A., YUILLE, J. C., & SMYTHE, P. C. Stimulus and response abstractness, imagery, and meaningfulness, and reported mediators in paired-associate learning. *Canadian Journal of Psychology,* 1966, **20,** 362–377.

RAVEN, J. C. *Guide to using the Coloured Progressive Matrices: Sets A, Ab, B.* London: H. K. Lewis, 1960.

ROHWER, W. D., JR. Social class differences in the role of linguistic structures in paired-associate learning: Elaboration and learning proficiency. Final report, United States Office of Education Basic Research Project No. 5-0605, Contract No. OE 6-10-273, November 1967.

ROHWER, W. D., JR. Mental mnemonics in early learning. *Teachers College Record,* 1968, **70,** 213–226. (a)

ROHWER, W. D., JR. Socioeconomic status, intelligence, and learning proficiency in children. Paper presented at the annual meeting of the American Psychological Association, San Francisco, September 1968. (b)

ROHWER, W. D., JR. Intelligence quotient versus learning quotient: Implications for elementary curricula. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, February 1969.

ROHWER, W. D., JR. Images and pictures in children's learning. *Psychological Bulletin,* 1970, **73,** 393–403. (a)

ROHWER, W. D., JR. Mental elaboration and proficient learning. *Minnesota Symposia on Child Psychology,* 1970, **4,** 220–260. (b)

ROHWER, W. D., JR., & LEVIN, J. R. Action, meaning and stimulus selection in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7,** 137–141.

ROHWER, W. D., JR., & LYNCH, S. Semantic constraint in paired-associate learning. *Journal of Educational Psychology,* 1966, **57,** 271–278.

ROHWER, W. D., JR., & LYNCH, S. Form class and intralist similarity in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6,** 551–554.

ROHWER, W. D., JR., & LYNCH, S. Retardation, school strata and learning proficiency. *American Journal of Mental Deficiency,* 1968, **73,** 91–96.

ROHWER, W. D., JR., LYNCH, S., LEVIN, J. R., & SUZUKI, N. Pictorial and verbal factors in the efficient learning of paired associates. *Journal of Educational Psychology,* 1967, **58,** 278–284.

ROHWER, W. D., JR., LYNCH, S., LEVIN, J. R., & SUZUKI, N. Grade level, school strata, and learning efficiency. *Journal of Educational Psychology,* 1968, **59,** 26–31.

ROHWER, W. D., JR., LYNCH, S., SUZUKI, N., & LEVIN, J. R. Verbal and pictorial facilitation of paired-associate learning. *Journal of Experimental Child Psychology,* 1967, **5,** 294–302.

ROHWER, W. D., JR., SHUELL, T. J., & LEVIN, J. R. Context effects in the initial storage and retrieval of noun pairs. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6,** 796–801.

RUNQUIST, W. N., & FARLEY, F. H. The use of mediators in the learning of verbal paired associates. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3,** 280–285.

SEMLER, I. J., & ISCOE, I. Comparative and developmental study of the learning abilities of Negro and white children under four conditions. *Journal of Educational Psychology,* 1963, **54,** 38–54.

STEVENSON, H. W., HALE, G. A., KLEIN, R. E., & MILLER, L. K. Interrelations and correlates in children's learning and problem solving. *Monograph of the Society for Research in Child Development,* 1968, 33 (No. 7).

WILSON, A. B. Social stratification and academic achievement. In A. H. Passow (Ed.), *Education in depressed areas.* New York: Teachers College, 1963.

(Received December 18, 1969)

Journal of Educational Psychology
1971, Vol. 62, No. 1, 15–16

REANALYSIS OF DATA FROM WORTHEN'S STUDY OF SEQUENCING IN TASK PRESENTATION[1]

BLAINE R. WORTHEN[2] AND JAMES R. COLLINS
Laboratory of Educational Research, University of Colorado

Two analyses used by Worthen are criticized and reanalyses of data are provided. When a covariate affected by the treatment was dropped and the unit of analysis changed from pupil scores to class mean, there were no treatment differences on any dependent variable, thus raising serious questions about Worthen's conclusion that "*S*s in Treatment D proved superior to *S*s in Treatment E on a majority of intertreatment comparisons."

The purpose of this note is to criticize and extend two analyses that were used in an earlier study (Worthen, 1968) and to caution against misinterpretation of the results of that study. There are two potential problems with the earlier analyses. First, the use of the pupil as the unit of analysis in both (*a*) the analysis of ratings of teachers by pupils and by observer-raters and (*b*) the analysis of the 10 outcome measures ignores possible interdependence of observations. A more conservative technique with respect to power in detecting treatment differences, the use of the class mean as the unit of analysis, was used in the reanalysis which is presented in Table 1.

Second, the legitimacy of using the Concept Knowledge posttest (which was influenced by the treatment) as a covariate in the analysis of other posttests and retention tests is questionable. Therefore, test scores from both transfer tests and both retention tests were reanalyzed using the design of the original study but deleting the Concept Knowledge posttest as a covariate. This reanalysis yielded no significant difference between treatments on any transfer or retention test.

Discussion

The reanalyses of measures of teaching behavior demonstrate that the teaching methods did systematically differ. However, when the outcome measures were reanalyzed using more rigorous analytic methods, no significant differences were found be-

TABLE 1

SUMMARY OF ANALYSES OF COVARIANCE OF CRITERION MEASURE POSTTEST SCORES AND RATINGS OF TEACHING BEHAVIOR BETWEEN TREATMENTS D AND E

Measure	df_1	df_2	*F*
Criterion measures			
Concept Knowledge Test (a)	1	7	3.54
Concept Retention Test 1 (b)	1	7	.832
Concept Retention Test 2 (b)	1	7	1.021
Concept Transfer Test (b)	1	7	.680
Negative Concept Transfer Test (b)	1	7	.39
Semantic Differential Attitude Scale (a)	1	7	.08
Statement Attitude Scale (a)	1	7	.46
Total Attitude Scale (a)	1	7	.81
Written Heuristic Transfer	1	7	2.02
Oral Heuristic Transfer	1	7	2.29
Measures of teaching behavior			
Observer ratings	1	7	804.00*
Pupil perception of teaching behavior	1	7	94.00*

Note.—In analyses of all criterion measures, IQ, arithmetic problem-solving and arithmetic computation were used as covariates. Pretest scores of the measure were also used as covariates in analyzing posttest scores of each measure marked "a" above. For retention and transfer tests, marked "b," concept knowledge pretest scores were used as additional covariates.

* $p < .001$.

---

[1] This note is a condensation of a paper (Worthen & Collins, 1970) that includes a detailed discussion of all theoretical and methodological issues involved in the reanalyses presented here.

[2] Requests for copies of this article should be sent to Blaine R. Worthen, Laboratory of Educational Research, University of Colorado, Boulder, Colorado 80302.

tween the treatments. In using class means as the unit of analysis, the lack of significant differences could be attributed to loss of power and it could be logically argued that if a "true" unit of analysis were available (see Peckham, Glass, & Hopkins, 1969), the results would doubtlessly fall somewhere between the extremes represented by these reanalyses and the earlier results presented by Worthen. However, when a questionable covariate used in the original study was deleted, a reanalysis based on the more powerful unit of analysis (individual subjects) employed by Worthen failed to yield significant differences between treatments. Viewed collectively, these reanalyses suggest that the results and conclusions related to the criterion measures in Worthen's study should be interpreted most cautiously.

REFERENCES

PECKHAM, P. D., GLASS, G. V., & HOPKINS, K. D. The experimental unit in statistical analysis. *Journal of Special Education*, 1969, **3**, 337–349.

WORTHEN, B. R. Discovery and expository task presentation in elementary mathematics. *Journal of Educational Psychology Monograph Supplement*, 1968, **59**(1, Pt. 2).

WORTHEN, B. R., & COLLINS, J. R. *Reanalyses and reinterpretation of task presentation data: A postscript to a previous study.* (Research Paper No. 44) Boulder: University of Colorado, Laboratory of Educational Research, 1970.

(Early publication received June 13, 1970)

Journal of Educational Psychology
1971, Vol 62, No. 1, 17–24

FACTORS AFFECTING MULTIPLE TRIAL FREE RECALL[1]

DAVID S. GORFEIN[2] AND CLAUDIA BLAIR
New College

Subjects were tested on a variety of material in multiple-trial free-recall paradigms, as well as over a number of single-trial free-recall lists. In addition all subjects participated in a paired-associate and short-term memory experiment. Correlational analysis showed relatively high correlations across materials of free-recall total score with a corresponding lack of correlation for subjective organization. A short-term memory score and a paired-associate learning score both correlated significantly with multiple-trial free-recall performance. A previous report that the correlation between free recall and paired associates was due to the fact that good paired-associate learners were good in subjective organization is not supported in this data. Significant practice effects were shown in several aspects of the total free-recall performance.

Melton (1967) in a symposium on individual differences in learning has suggested that in order to refine process theories of learning an analysis of individual differences in these processes would be useful. Underwood (1967) has emphasized the importance of examining the relationships between rote verbal learning tasks in the development of a theory of memory. Despite these calls for action comparatively little has been done in either the area of individual differences in memory or in the examination of the relationships between tasks. Whimbey and his colleagues (Ryan & Whimbey, 1968; Whimbey & Leiblum, 1967) have begun extensive investigations of the digit-span task and its relationships to other situations. In the area of concern of the current paper, Earhard (1967) and Earhard and Endicott (1969) have studied the effects of using subject groups differing in subjective organization ability (Tulving, 1962) on performance in several tasks. They found high-subjective-organization subjects showing better serial recall than low-subjective-organization subjects and have recently found those high in subjective organization to be better in paired-associate learning. The current authors have been engaged in the study of individual differences in free recall and their relationships to performances in other tasks. Gorfein and Bennett (1967) reported significant correlations between trial performance in a multiple-trial free-recall task and short-term memory processes for CCCs in a Peterson and Peterson (1959) paradigm. In a study aimed at examining the generality of free recall (Gorfein, Blair, & O'Neill, 1969; Gorfein, Blair, & Rowland, 1968a, 1968b), it was demonstrated that a fair correlation of performance measures (such as rate of item gain, the relative gain of score and individual trial performance) is obtained across a variety of materials (CCCs, CVCs, words, and categorized words). In unpublished results from that study it was again confirmed that short-term memory correlates significantly with trial scores from all materials. However, for this data it was observed that there were not significant correlations of subjective organization scores (Fagan, 1968). Further, subjective organization for the CCCs failed to correlate with performance on CCCs.

The current study is designed to explicate further the relationship between subjective organization and meaningfulness and between performances in free recall of a variety of materials. At the same time the relationship of free-recall performance to performance in paired associates and short-term memory were investigated.

---

[1] The analysis of the data and writing of this paper were partially supported by the Human Performance Center, University of Michigan, Arthur W. Melton, Director.

Part of the research reported represents work submitted by the junior author in fulfillment of the senior thesis requirement of New College.

[2] Requests for reprints should be sent to David S. Gorfein, Natural Science Division, New College, Sarasota, Florida 33578.

METHOD

Subjects

Thirty-two volunteer first-year New College students (15 men and 17 women) participated in all parts of this study and were paid for their time.

Procedure

The major focus of this study was to investigate the relationships between the multiple-trial free recall of a variety of material. Prior to the collection of the multiple-trial data all subjects participated in the collection of single-trial free-recall data[3] in a single session. For single-trial free recall, 24 lists of 24 words were randomly drawn from the A and AA words in the Thorndike-Lorge (1944) list. Words were randomly assigned to lists with the restriction that there were no obvious intralist associations and no two words occurring consecutively in a list began with the same letter. Items were presented visually, using a Kodak Carousel projector and a 1-second rate. Three minutes were given after each trial for written free recall.

Multiple-trial free-recall sessions consisted of seven presentations of a given list at a 1-second rate with 4 minutes allowed after each trial on the 40-item list for written free recall. Six sets of multiple-trial free-recall items were used. Four trigram lists were employed balanced for first letter repetitions and meeting a criterion of no item in any list being the reverse of any item in a list and no first two letter combinations being repeated. The four trigram lists consisted of Witmer (1935) 67% association value CCCs and three lists of CVCs, one each of low (CVC_l), medium (CVC_m) and high (CVC_h) meaningfulness (Noble, 1961). Finally, two word lists (Word List 1 and Word List 2) matched for Thorndike-Lorge (1944) frequency and having a mean frequency of 7.35 words per million were employed in the study. Only a single list was presented in a session and a minimum of 24 hours elapsed between sessions. Half of the group, selected at random, received the lists in the order of CVC_h, CVC_l, Word List 1, CCC, CVC_m, and Word List 2, while the other half of the group received the lists in reverse order. In all sessions the items were presented in random order at the first trial and then shuffled and returned to the projector before the next trial.

Subsequent to the completion of the free-recall portion of the experiment all subjects were individually tested on a 15-item paired-associate list. Four random orders of a list using the numbers 1 to 15 as stimuli and medium-meaningful CVCs (Noble, 1961) as responses were used. The subjects were given paired-associate instructions and required to learn the list to the criterion of one perfect trial at a 2:2 rate. The paired-associate measure was the number of trials to criterion, hence high score represents a slow learner. Twenty-four hours later short-term memory was tested. A standard Peterson and Peterson (1959) paradigm was used in which, following 18 practice items, each subject was individually tested on 15 trials in which they were required to recall consonant quadrigrams. Each subject was assigned a short-term memory score based on the number of quadrigrams he remembered.

RESULTS AND DISCUSSION

Meaningfulness and Performance in Free Recall

Figure 1 shows the growth in performance over trials of all six multitrial lists. The high-meaningful CVCs (which were all three-letter words) and the two word lists do not differ significantly in performance. As a group, the high-meaningful material is superior to all other material. The medium-meaningful list is superior to the low-meaningful materials which do not differ significantly.

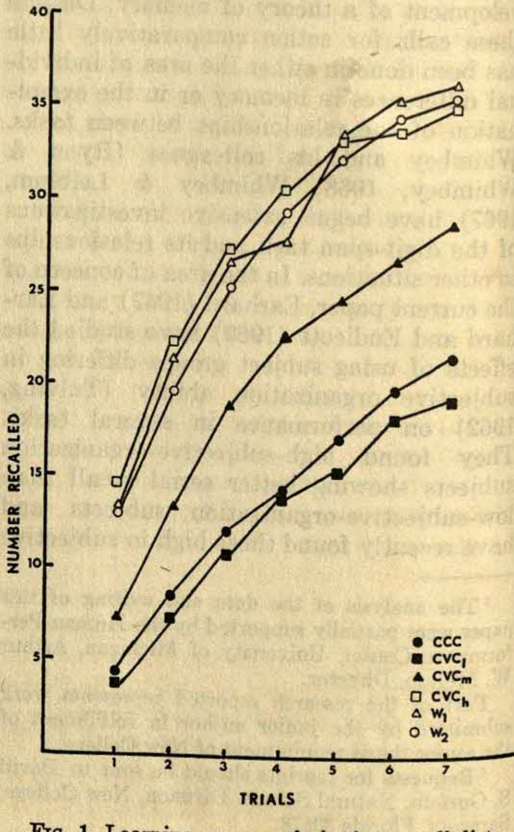


FIG. 1. Learning curves of six free-recall lists.

---

[3] The single-trial free-recall, short-term memory and paired-associate data were collected by Christopher Arbak and Diana Graves, as part of an independent study project.

The authors had anticipated the general form of the results on trial performance, that is, high-meaningful materials being learned faster than the medium- and low-meaningful materials. It was interesting to note that there was no significant difference in the learning rates of the CVC and CCC groups. No norms exist for the equating of the meaningfulness of CCCs and CVCs but it is certainly the case (see Underwood & Schulz, 1960) that the low-meaningful CVCs are easier to pronounce than the CCCs. This being the case one would either have to conclude that pronounceability played no role in free-recall performance or that the pronounceability differences were offset by differences in the opposite direction in meaningfulness.

Practice Effects in Free Recall

Although the study was not intended as a study of practice effects in free recall, the lack of difference in the learning curves among the three high-meaningful materials and their locus in the first, middle, and last positions of the counterbalanced list order makes it possible to do analyses of the practice effects. An analysis of average trial performance (total number of items remembered over seven trials divided by seven) shows an average of 25.96 items correctly remembered for the list presented first (no prior multiple-trial free-recall tests), 28.41 for the middle list, and 30.37 for the final list ($F = 27.35$, $df = 2/62$, $p < .001$). Tulving (personal communication, March 1968) has provided a measure of bidirectional subjective organization based on his (1962) undirectional measure of subjective organization. For each subject over the seven trials on each free-recall list the authors computed this bidirectional subjective organization score (SO2). On the first high-meaningful list, subjects had a mean of .269, on the middle list a mean of .333, and on the final list a mean of .347. There was a significant practice effect across the three high-meaningful materials ($F = 116.49$, $df = 2/62$, $p < .001$). A correlation between gain in SO2 and gain in performance from the first to last list was significant ($r = .54$, $p < .01$).

The practice effects observed in the performance on the high-meaningful materials support the findings of Mayhew (1967) who also found a significant effect of practice on subjective organization and on total score. In Mayhew's results the effect on performance fell short of statistical significance possibly due to the very small sample ($n = 12$) employed in each of his conditions.

TABLE 1

CORRELATIONS OF FREE-RECALL TOTAL SCORES

List	Word List 2	$\mathrm{CVC_h}$	$\mathrm{CVC_m}$	$\mathrm{CVC_l}$	CCC	Total single trial
Word List 1	.77	.66	.75	.60	.67	.66
Word List 2	—	.63	.73	.61	.69	.66
CVC (high meaningful)		—	.65	.66	.62	.71
CVC (medium meaningful)			—	.76	.72	.72
CVC (low meaningful)				—	.67	.60
CCC					—	.53

Note.—For $r \geq .33$, $p < .05$; $r \geq .41$, $p < .01$.

Since there were highly significant practice effects for the purposes of all the correlational analyses that follow all scores were converted to a standard score within the subgroup (standard score $= 100Z + 500$).

Consistency of Individual Differences in Free Recall

Table 1 reports the intercorrelations of total performance (total number of items remembered over the seven trials by each subject) of the six multitrial free-recall lists. In addition it shows the correlation with each subject's total single-trial performance on the 24 single-trial lists. Correlations across multiple-trial materials are reasonably high, accounting for about 40%–50% of the total variance in multiple-trial free-recall performance. These correlations are particularly notable if one considers the correlation between the two matched word lists ($r_{w1,w2} = .77$) as an estimate (although confounded by practice effects) of test-retest reliability of free-recall performance. It is certainly clear that it is likely that the same factors to a large extent enter into free-recall performance over a wide variety of verbal materials.

Table 2 reports the intercorrelations of the rate of gain of new items. The score used for correlational purposes was the me-

TABLE 2

CORRELATIONS INVOLVING RATE OF GAIN OF NEW ITEMS

List	Word List 2	$\mathrm{CVC_h}$	$\mathrm{CVC_m}$	$\mathrm{CVC_l}$	CCC	Total score (App.)[a]
Word List 1	.73	.64	.46	.56	.60	.87
Word List 2	—	.57	.57	.61	.60	.90
CVC (high meaningful)		—	.42	.43	.62	.85
CVC (medium (meaningful)			—	.69	.56	.84
CVC (low meaningful)				—	.58	.84
CCC					—	.90

Note.—For $r \geq .33$, $p < .05$; $r \geq .41$, $p < .01$.

[a] App. refers to the appropriate score subscripted, that is, that derived from the material being correlated.

dian rate of gain over the first four trials where the rate measure is equal to number of new items gained on a trial (words on the list not previously reported by the subject on an earlier trial) over 40 (the total number of items) minus the number of items already gained. The gain measure is therefore a relative gain measure. The correlations in this table are on the average smaller than those of the total free-recall scores. The rate of gain of new items does correlate significantly with total score, in fact showing a minimum of 70% of common variance with the total scores. Mean relative gain in items for materials was an average of 44.09 for the three high-meaningful lists, 27.30 for medium-meaningful lists, and 15.35 for the two low-meaningful materials.

For each subject a score was computed on each material which represented the retention of newly gained items. To derive this score the number of items gained on Trial *N* and retained to trial *N* + 1 were summed over the seven trials and divided by the total of new items gained in the course of Trials 1–6 (Trial 7 gain is not used because there is no measure of retention to a succeeding trial.) Mean retention score for the two low-meaningful materials averaged 58%, for the medium-meaningful trigrams the mean was 66%, and for the high (averaged over the three materials) the retention score had a mean of 78%. There was a practice effect shown in the high materials with the retention percentage going from approximately 73% on the first list to 81% on the middle lists with no further improvement on the final lists. The intercorrelations of the retention measures are considerably smaller than either the total score or gain measures. Four of the correlations fall short of statistical significance. All of these involve the two low-meaningful materials. Examination of the relationships among the SO2 scores provide further support to the conclusion reached in the previous paper (Gorfein, Blair, & O'Neill, 1969) that subjective organization is not a reliable individual difference variable across a range of materials. Mean SO2 scores were .316 for the high-meaningful lists, .261 for medium meaningfulness, .226 for low meaningfulness, and .263 for CCCs.

The range of subjective organization scores across materials is quite small. It would seem that such a small range of subjective organization scores could hardly account for the large differences in performance observed across materials. It is per-

TABLE 3

CORRELATIONS OF RETENTION OF NEW ITEMS GAINED

List	Word List 2	$\mathrm{CVC_h}$	$\mathrm{CVC_m}$	$\mathrm{CVC_l}$	CCC	Total score (App.)[a]	Rate of gain
Word Lits 1	.49	.40	.49	.31	.62	.84	.51
Word List 2	—	.45	.44	.16	.31	.69	.57
CVC (high meaningful)		—	.43	.46	.40	.75	.66
CVC (medium meaningful)			—	.42	.43	.76	.61
CVC (low meaningful)				—	.21	.65	.41
CCC					—	.76	.62

Note.—For $r \geq .33$, $p < .05$; $r \geq .41$, $p < .01$.

[a] App. refers to the appropriate score subscripted, that is, that derived from the material being correlated.

haps the case that the subjective organization score for the low-meaningful materials is artificially inflated by the fact that some of the poorest subjects tend to remember the same items from trial to trial and achieve very high SO2 scores. This interpretation is supported by the observed lack of correlation of the subjective organization measure with any of the free-recall measures for low-meaningful materials (see Tables 5, 6, and 7).

Of far greater significance for the development of theoretical models of free recall are the results obtained examining the intercorrelation of the free-recall variables. The failure to obtain large and significant correlations of the subjective organization scores (even just among the high-meaningful materials) casts some doubt on subjective organization as a theoretical process variable (cf. Melton, 1967). It would certainly seem that this correlation is not significantly attenuated by practice effects (it is computed on standard scores) and it is usually the case with practice effects that subjects maintain their relative position on a variable with practice. The significant correlation obtained between the gain in SO2 from the first to last list and the gain in average trial performance from the first to last list do suggest that subjective organization is an important factor in the learning of high-meaningful materials. The relatively high correlation across materials of free-recall total score with the corresponding lack of correlation for subjective organization suggest that a fair amount of the variability in free-recall performance is due to other factors common to the free-recall situation.

TABLE 4
CORRELATIONS OF SUBJECTIVE ORGANIZATION (SO2) SCORES

List	Word List 2	CVC_h	CVC_m	CVC_l	CCC
Word List 1	.40	.06	.10	.20	.02
Word List 2		.33	.37	.02	.20
CVC (high meaningful)			.28	.36	−.31
CVC (medium meaningful)				−.04	.05
CVC (low meaningful)					.11

Note.—For $r \geq .33$, $p < .05$; $r \geq .41$, $p < .01$.

Relationship of Free-Recall Paired-Associate Learning and Short-Term Memory

In order to examine the relationship between the three verbal tasks, the short-term memory and paired-associate tasks have been treated as predictors of total score in free recall. Table 5 reports the correlations of these tasks as well as the correlation of subjective organization with total score in free recall. Subjective organization correlates significantly with performance for all but two low-meaningful materials. There are, of course, no measures of subjective organization for single-trial free recall (see Gorfein, Bennett, Arbak, & Graves, 1969, for a comparable analysis of the single-trial situation). As was observed in the previous studies, short-term memory correlates significantly with free-recall performance although correlation with Word List 2 falls

TABLE 5
PREDICTORS OF TOTAL SCORE

Variable	Predicted material						Total single trial
	Word List 1	Word List 2	CVC_h	CVC_m	CVC_l	CCC	
SO2 (App.)[a]	.67	.70	.43	.53	.16	.04	—
Short-term memory	.43	.28	.47	.46	.53	.46	.32
Paired associates	−.42	−.51	−.38	−.61	−.57	−.53	−.48
Multiples							
3 variables	.77	.82	.68	.78	.72	.67	—
Short-term memory and paired associates	.54	.54	.54	.69	.70	.63	.53

Note.—For $r \geq .33$, $p < .05$; $r \geq .41$, $p < .01$.
[a] App. refers to the appropriate score subscripted, that is, that derived from the material being correlated.

6. 3. 81

just short of statistical significance. It should be noted that short-term memory does not correlate significantly with subjective organization, the six correlations computed ranging from 0 to .30 with the median about .10. Paired-associate learning correlates significantly with all free-recall performances. For the low-meaningful material (medium CVCs to CCCs) the correlations indicate a minimum of 28% common variance between the two tasks. It should be noted here that the response terms for paired-associate tasks were medium-meaningful CVCs. The correlation between paired-associate learning and subjective organization scores ranged from +.19 to −.25 with a median correlation of −.09 (none of these correlations reaching statistical significance). The correlation of paired associates and short-term memory was −.24.

It seemed reasonable to assess how much of the variability in total free-recall performance could be accounted for by treating these three correlates (SO2, short-term memory, and paired-associates) as predictors of total performance in a multiple correlation. Using a Doolittle solution (McNemar, 1962) we found that the multiples could account for 45%–67% of the variance in the total score performance. Looking at the two independent variables (short-term memory and paired associates) alone we find that their multiple can account for 28%–49% of the variability in free-recall total performance.

In further analysis we attempt to see the relationships between the three predictor variables and aspects of the free-recall performance. In Table 6 the correlations involving rate of gain of new items are reported. Subjective organization fails to correlate as well with the rate measure as it did with total performance, with three of the correlations, those for CVC_h, CVC_l, and CCCs falling short of statistical significance. Short-term memory shows about comparable correlation with the exception of a large gain in the prediction of medium-meaningful CVCs. Paired associates show slightly smaller correlations for the most part. The multiple correlations employing three predictor variables are in general smaller than those predicting total score, with a single exception of the prediction of the rate of gain of items in the medium CVC list. In this situation the multiple correlation accounts for about 67% of the variance in medium-meaningful CVCs. Most of the changes in multiple correlation were produced by the smaller correlation of SO2 with the gain measure. This is clear in looking at the prediction of the two variables, short-term memory and paired associates, of the rate of gain measure. They predict the rate measure about as well as they predicted total score.

As would have been anticipated from the examination of the relationships in Table 3, the correlations between the predictor variables and the retention scores are on the average smaller and a good deal less homogenous. Nevertheless, the three-variable multiple does account for from 25%–62% of the variance in the score and the two-varia-

TABLE 6

PREDICTORS OF RATE OF GAIN OF NEW ITEMS

Variable	Predicted material					
	Word List 1	Word List 2	CVC_h	CVC_m	CVC_l	CCC
SO2 (App.)[a]	.45	.58	.15	.44	.21	−.13
Short-term memory	.42	.44	.45	.64	.55	.42
Paired associates	−.43	−.44	−.36	−.61	−.44	−.61
Multiples						
3 variables	.62	.76	.54	.82	.66	.67
Short-term memory paired associates	.54	.56	.52	.79	.63	.67

Note.—For $r \geq .33$, $p < .05$; $r \geq .41$, $p < .01$.

[a] App. refers to the appropriate score subscripted, that is, that derived from the material being correlated.

TABLE 7

PREDICTORS OF RETENTION SCORES

Variable	Predicted material					
	Word List 1	Word List 2	CVC_h	CVC_m	CVC_l	CCC
SO2 (App.)	.73	.72	.37	.47	.21	.20
Short-term memory	.32	0	.43	.19	.29	.50
Paired-associates	−.39	−.46	−.30	−.56	−.40	−.26
Multiples						
3 variables	.78	.79	.59	.68	.50	.62
Short-term memory and paired associates	.45	.47	.48	.56	.45	.52

Note.—For $r \geq .33, p < .05; r \geq .41, p < .01$.

[a] App. refers to the appropriate score subscripted, that is, that derived from the material being correlated.

ble multiples from 20%–31% of the variability.

The correlations between short-term memory in the Peterson and Peterson situation and aspects of the free-recall performance that we have now observed in a number of studies suggest that one of the factors determining the commonality of free-recall performance may be the ability to encode items individually and uniquely enough to be retrievable. This conclusion is based on an assumption that individual differences in short-term memory are produced by subjects varying in the ability to encode items so as to overcome proactive inhibition (Wickens, 1968, has documented the case for the role of proactive inhibition in short-term memory). Further, an analysis of the correlation of short-term memory and portions of the single-trial free-recall curve (Gorfein et al., 1969) indicates that short-term memory has its primary influence in the primary and middle portions of the curve. One would anticipate that the ability to uniquely encode items would have its greatest influence here (cf. Postman & Phillips, 1965).

An interpretation of the correlation between the paired-associate measure and the free-recall scores does not seem obvious. Earhard and Endicott (1969) had suggested that paired-associate learning would represent the ability of subjects to associate items and would be comparable to subjective organization. However the nonsignificant correlations obtained between paired-associate learning and any of the SO2 scores strongly argue against this factor. In addition we must note the failure of the paired-associate measure to correlate very highly with the short-term memory score indicating that the observed correlation is probably not due to a general verbal learning ability.

In taking an overview of the data reported we are impressed by the high correlations between the total free-recall performances as well as the consistent correlations with paired-associate learning and short-term memory. These correlations occur despite the wide variety of materials. Their magnitudes are quite consistent despite a fair amount of unreliability in the measures. The data suggest that a theoretical model of free recall can be built with minimal attention to Subject × Material interactions.

REFERENCES

EARHARD, M. Subjective organization and list organization as determinants of free recall and serial recall memorization. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6,** 501–507.

EARHARD, M., & ENDICOTT, O. Why are there individual differences in subjective organization during free-recall memorization? *Journal of Verbal Learning and Verbal Behavior,* 1969, **8,** 316–319.

FAGAN, J. F., III. Measuring verbal recall. II: The ITR score expressed as a ratio. *Psychonomic Science,* 1968, **11,** 6.

GORFEIN, D. S., & BENNETT, R. W. Analysis of free-recall over multiple trials. Paper presented at the annual meeting of the Psychonomic Society, Chicago, October 1967.

GORFEIN, D. S., BENNETT, R. W., ARBAK, C., & GRAVES, D. Individual differences in the serial position curve of free-recall. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8,** 769–772.

GORFEIN, D. S., BLAIR, C., & O'NEIL, C. R. A reanalysis of the generality of free-recall: I. Subjective organization as an ability factor. *Psychonomic Science,* 1969, **17,** 110.

GORFEIN, D. S., BLAIR, C., & ROWLAND, C. The generality of free-recall: I. Subjective organization as an ability factor. *Psychonomic Science,* 1968, **11,** 279–280. (a)

GORFEIN, D. S., BLAIR, C., & ROWLAND, C. The generality of free-recall: II. Performance as a function of type of material. *Psychonomic Science,* 1968, **13,** 111–112. (b)

MAYHEW, A. J. Interlist changes in subjective organization during free-recall learning. *Journal of Experimental Psychology,* 1967, **74,** 425–430.

MCNEMAR, Q. Psychological statistics. (3rd ed.) New York: Wiley, 1962.

MELTON, A. W. Individual differences and theoretical process variables. In R. M. Gagné (Ed.), *Learning and individual differences.* Columbus, Ohio: Charles E. Merrill, 1967.

NOBLE, C. E. Measurement of association value (a), rated associations (a′) and scaled meaningfulness (m′) for the 2100 CVC combinations of the English alphabet. *Psychological Reports,* 1961, **8,** 487–521.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Educational Psychology,* 1959, **58,** 193–198.

POSTMAN, L., & PHILLIPS, L. Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology,* 1965, **17,** 132–138.

RYAN, S. F., & WHIMBEY, A. E. STM abilities in LTM tasks. *Psychonomic Science,* 1968, **10,** 297–298.

THORNDIKE, E. L., & LORGE, I. *The teachers word-book of 30,000 words.* New York: Teachers College, 1944.

TULVING, E. Subjective organization of free-recall of unrelated words. *Psychological Review,* 1962, **69,** 344–454.

UNDERWOOD, B. J. The representativeness of rote verbal learning. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1967.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago, Ill.: Lippincott, 1960.

WHIMBEY, A. E., & LEIBLUM, S. L. Individual differences in memory span with and without activity intervening between presentation and recall. *Journal of Educational Psychology,* 1967, **58,** 311–314.

WICKENS, D. Encoding categories in short-term memory. Presidential address to Division 3 of the American Psychological Association, San Francisco, September 1968.

WITMER, L. R. The association value of three-place consonant syllables. *Journal of Genetic Psychology,* 1935, **47,** 337–360.

(Received August 24, 1969)

Journal of Educational Psychology
1971, Vol 62, No. 1, 25-30

SENTENCE STRUCTURE AND RETENTION IN GOOD AND POOR READERS[1]

RHONA WEINSTEIN
Yale University

AND

M. SAM RABINOVITCH[2]
McGill University

The purpose of the experiment was to discover whether syntactic structure facilitates recall in good readers, and whether this effect exists in children who are poor readers. A paired-associate task equated the two groups on their ability to associate simple, familiar words. Each child learned (with a tape recorder) four lists composed of nonsense elements and grammatical markers; two of which were syntactically structured, the other two unstructured. The good readers learned the structured lists more rapidly than the unstructured lists. The poor readers learned both kinds of lists with equal difficulty. There was no difference between the good reader's and the poor reader's ability to retain the unstructured material. Hence, the locus of the facilitation effect lies in the syntactic cues, implicit in the structured list.

Some children of normal intelligence encounter multiple difficulties in learning to read and write. An essential component of reading retardation or dyslexia is difficulty with sequential ordering. Orton (1937) proposed that the interfering factor in the recognition or recall of a word is a failure to reconstruct the exact order of its constituent letters. Lashley's (1951) discussion of the serial order phenomenon makes it clear that the individual items of a temporal series, for example, words, do not in themselves have an intrinsic valence. He argues that the order is imposed by broader relations, a set of rules, commonly called syntax. Lashley (1951) suggests that "... this is the essential problem of serial order; the existence of generalized schemata of action which determine the sequence of specific acts [p. 122]."

The generalized schemata, in linguistic terms, are commonly referred to as grammatical markers (Fries, 1962), devices that signal the structural meaning of a word. Any word, in a complete utterance, when given a specific identifying marker, will be tagged as a member of one of four major form classes (noun, verb, adjective, and adverb). Fries' (1962) classification of signals consist of (*a*) the contrastive arrangement of form classes, bound morphemes, (*b*) function words, and (*c*) patterns of intonation. These signals cannot be violated or structural ambiguities will result. In language learning, grammatical markers impose a predictable order on the reproduction of the elements composing the sentence. By investigating recall of a span of verbal material in children, one could determine how effectively or ineffectively these youngsters use the information available in the decoding of the grammatical markers.

Epstein (1961, 1962) has reported that in adults the recall of syntactically structured lists of nonsense syllables is superior to the recall of unstructured lists of nonsense syllables. The lists he used included three basic components: (*a*) English function words, (*b*) nonsense syllable stems (to control the semantic variable), and (*c*) English bound morphemes. Epstein's experiment investigated three types of nonsense sentences. The structured list could be described as a grammatical "pseudo-English" sentence, since if its nonsense stems were replaced by English stems, a grammatical sentence would result. The unstructured list contained the same items as the structured list, but in a scrambled order. It is unstructured in the sense that word order is ungrammatical. The third list was identical to the structured list except that the bound morphemes had been deleted. This latter type of sentence is not ungrammatical. However, the number of syntactic cues available to the reader or listener is reduced. Epstein (1961, 1962) found that structured lists, with and without bound morphemes, are learned in fewer trials than unstructured lists, and the addition of bound morphemes further facilitates recall. The function words (e.g., the) and the bound morphemes (e.g., -ed) serve as grammatical markers which associate a form class to each nonsense stem. If a listener has perceived the syntactic structure of the list, he need not store information about the order in which the items appear. This information is explicit in the markers themselves. Recall may thus be related to the number of syntactic cues present in the list.

---

[1] The present paper, based on the senior author's master's thesis at McGill University, is a revised and expanded version of one presented at the 1967 Conference on Reading Difficulties in Children held at the Douglas Hospital, Montreal, Canada, the proceedings of which were published in The Canadian Psychologist, 1968, Volume 9.

Thanks are due the members of the Protestant School Board of Greater Montreal for their assistance.

[2] Requests for reprints should be sent to M. Sam Rabinovitch, Department of Psychology, McGill University, Montreal, Canada.

In Epstein's (1961, 1962) experiment, subjects were asked to recall the items in order. Forster (1966) proposed that the facilitation effect apparent in the structured lists was due to the fact that the subject is not required to store information concerning item order. He tested the prediction that if subjects were not obligated to recall the items in order, the facilitation effect would disappear. However, a facilitation effect did appear; structured lists with and without bound morphemes were recalled more efficiently than unstructured lists, even under conditions of nonordered recall. The crucial variables in these studies, then, are the syntactic cues within the list, whether recall of the order of items is insisted upon or not.

In the present study, it was postulated that both good and poor readers would be able to associate random words with the same facility. To test this, a paired-associate task was given to all subjects. However, in a grammatically structured string of words, the most efficient way to remember and reproduce the order of items is to use the information imparted by the syntactic markers. The prediction was that poor readers would not respond to the syntactic cues and therefore would not be able to recall the serial order of the items any easier with the aid of markers than without them.

In order to test these hypotheses with children, it was necessary to generate a simpler set of lists than those used by Epstein (1961, 1962) and Forster (1966). Appropriate sentence frames were taken from children's readers and a new set of nonsense stems were generated using Forster's (1966) rules. A pilot study was done to determine the difficulty of the lists prepared, difficulty in terms of length or number of items to be learned. Previous experiments (Epstein, 1961, 1962; Forster, 1966) tested recall under the conditions of reading the material and writing the reproduction. Since poor readers are at a great disadvantage under such conditions, reading was eliminated from the experiment entirely. A tape recorder was used throughout the experiment, and the children were asked to repeat exactly what they heard.

METHOD

Subjects

Fourth-grade children from three Montreal public schools were the sample pool from which the subjects were chosen. They all spoke English as their native language and were all of approximately the same socioeconomic level (middle class).

These children were given the Gates Reading Test (Gates, 1958) in the second month of the school year. Reading ability level was determined by the average of the three subtests, speed and accuracy, vocabulary, and comprehension. All the children in the upper third of the distribution of scores on the Gates Reading Test were selected for the group of good readers. Children whose scores fell at the 50th percentile or lower defined approximately the lower third of the Grade 4 reading population and formed the group of poor readers.

Since reading ability correlates with IQ scores (Muehl & Kremenak, 1966), an effort was made to equate the two groups as closely as possible for IQ as measured by the Henmon-Nelson Tests of Mental Ability (Henmon & Nelson, 1958). The mean IQ for the good readers was 112.46 with a range of 96 to 127, for the poor readers, 103.80 with a range of 95 to 116. This IQ difference is statistically significant at the .001 level ($t = 3.59$), and this made it necessary to use an analysis of covariance approach to the data analysis.

Ten children were withdrawn at the outset of the testing sessions due to speech problems, emotional reactions to the material, and initial mechanical difficulties with the tape. In all, 41 children were included in the sample, 26 good readers and 15 poor readers.

At the time of testing, the children ranged in age from 8.91 to 11.33 years. Mean age for the good readers was 10.09 and for the poor readers, 10.04. Mean reading level significantly differentiated the two groups, Grade 5.70 as compared to Grade 4.04, ($t = 11.84, p < .001$).

Preparation of Lists

1. *Nonsense syllable stems.* With the aid of a table of bigram frequencies in English (Underwood & Schulz, 1960), separate strings of 60 letters each were prepared using a table for random numbers. Each string was cut into segments. The length of each segment was determined by English word stems which were randomly selected from the Thorndike-Lorge (1944) word list (Words 1 to 500). The common words that children use comprise this list. As an example, if the first four English word stems had six, four, three, and five letters, respectively, then the string of bigrams was segmented after the sixth, tenth, thirteenth, and eighteenth letters. This procedure generated potential nonsense-syllable stems. All syllables that were actual English words were discarded, as were syllables that violated English orthographic rules, and were judged not pronounceable at the Grade 4 level. The stems used in the experiment were: (*a*) sivol, veg, hanash, zalf; (*b*) doo, swad, gozur, kaffal; (*c*) onuc, hend, tik, parf, imi; (*d*) rak, ibnu, lurm, wab, elir.

2. *Construction of lists.* Four English frames were selected from a Grade 3 reader. They were as follows:

(*a*) When they __ed the __, they __ed __ly.
(*b*) The __s __ed on the __quite __ly.
(*c*) __s __ed __as if __ing to __.
(*d*) All the __ __ __ed and __ed __ly.

The structured lists were constructed by inserting the stems in the available positions in the frames. Four unstructured lists were prepared by randomly rearanging the order of the items in each structured list.

Procedure

TABLE 1

CONTENT OF THE FOUR ORDERS

Order	Code	Sentence
1	A1S	When they sivoled the veg, they hanashed zalfly
	B1S	Onucs hended tik, as if parfing to imi
	A2R	Quite the swaded the, on kaffally doos gozur
	B2R	And all rak elirly, wabed ibnu the lurmed
2	A1R	Zalfly they when, veg the hanashed, sivoled they
	B1R	Imi onucs, if tik, to hended parfing as
	A2S	The doos swaded on the gozur quite kaffally
	B2S	All the rak ibnu lurmed and wabed elirly
3	A2S	The doos swaded on the gozur quite kaffally
	B2S	All the rak ibnu lurmed and wabed elirly
	A1R	Zalfly they when, veg the hanashed, sivoled they
	B1R	Imi onucs, if tik, to hended parfing as
	A2R	Quite the swaded the, on kaffally doos gozur
	B2R	And all rak elirly, wabed ibnu the lurmed
	A1S	When they sivoled the veg, they hanashed zalfly
	B1S	Onucs hended tik, as if parfing to imi

Note.—Abbreviated: A = Type A, B = Type B, 1 = first exemplar of type, 2 = second examplar of type, S = structured, R = unstructured.

Under conditions of ordered recall, the children were asked to learn four sentences, two of which were structured, and two, unstructured. Four orders were prepared for the experiment, and they are presented in Table 1. The first two frames were designated as Type A since they were matched as to number of function words, number of bound morphemes, and number of stems. The third and fourth frames constitute Type B and were matched with each other, according to the same criteria. Each child would learn one A and one B frame in its structured form and the other A and other B frame in its unstructured form. Having four distinct orders (i.e., the material in each was different) made it possible for half of the children to learn each sentence in its structured form, while the other half learned the same sentence in its unstructured form. The design also allowed an equal number of children to learn structured lists, followed by the unstructured lists; while the other half learned the unstructured lists first. The assignment of each child to one of the four orders was determined by order of appearance. The orders were constructed so as to counteract the possible effects of (*a*) one list being equivalent to another in number of function words, number of bound morphemes, and number and length of nonsense stems, yet learned more easily, and of (*b*) learning either structured or unstructured lists first, in terms of producing a set.

Each child was tested individually, for approximately 1 hour. Each order was recorded on tape by a male voice. The child listened to the material through earphones having been instructed verbally by the examiner that he would hear a funny sentence. He was asked to say aloud exactly what he heard in exactly the way he heard it. Immediately after the approximately 10-second learning period, the child was given as long as he needed to reproduce the list verbally. If incorrect, the material was presented again, followed by another test. This procedure was repeated until three reproductions, correct in terms of item order, were obtained. All four lists were learned by each child in this way.

Prior to the list-learning period of the experiment, each child, both good reader and poor reader, was given a paired-associate task containing six pairs of words. The words were chosen randomly from the Thorndike-Lorge (1944) word list (Words 1–500). The selection criteria were that (*a*) each word contain four letters and (*b*) no two words of a pair begin with the same letter. The six pairs of words are presented in Table 2. The procedure was identical to that for the lists. The child listened through earphones to the recording of the pairs. The learning period took approximately 45

TABLE 2

PAIRED ASSOCIATES

Pairs	
food	soft
dark	pass
half	them
been	milk
call	near
last	what

seconds and the interval between training and test, approximately 10 seconds. The child's verbal responses were checked—if incorrect, all the pairs were replayed until the subject answered all the pairs correctly on three successive trials.

RESULTS

The mean number of trials required by each subject to reach the criterion of three correct verbal reproductions was computed for the paired-associate task, the structured lists, and the unstructured lists. Before this could be done for the nonsense material, it was essential to define empirically a correct reproduction. This was made difficult by the fact that the learning trials were in the auditory rather than the visual modality. Both the strangeness of the material, a combination of familiar with unfamiliar, and the human voice imperfections in the tape recording left room for a variety of minor misperceptions.

It was necessary to arrive at a set of workable rules that would define an acceptable response. The criteria that were held as most important were that the item order of the list must remain intact, and that all bound morpheme endings and function words must be pronounced correctly. If on any trial, either of these rules were violated, the trial was marked as incorrect. All of the children's less than perfect reproductions of the nonsense stems were analyzed in three ways: (*a*) in terms of the frequency of occurrence of that approximation in good and poor readers; (*b*) in terms of the similarity of the given response to the correct response, for example, "ini" instead of "imi"; (*c*) in terms of the frequency of mispronunciation of the initial consonant. Many children began the word "veg" with a letter other than "v," therefore, all the initial consonants used were considered acceptable. Whereas, in the case of "tik," only one child began the word with other than a "t"; hence, that approximation was scored as an error. All approximated responses were empirically evaluated in this way until a table of acceptable approximations was generated to be used in scoring all the responses.

On the paired-associate task, the mean number of trials required to reach criterion was 14.23 for the good readers and 17.00 for the poor readers. This is not a significant difference ($t = 1.25$), suggesting that the good and poor readers can learn to associate simple familiar words with the same ease.

A comparison of trials to criterion for structured lists and unstructured lists in both groups is presented in Table 3.

Good readers learned structured lists in fewer trials than unstructured lists ($t = 7.67$, $p < .001$). That is, the syntactic cues facilitated recall for the good readers. In the group of poor readers, the mean number of trials for structured lists compared with the mean for unstructured lists is not significant ($t = 1.72$). The facilitative effect on retention associated with syntactic structure in good readers is not evident in the poor readers.

Only the presence of syntactic structure distinguishes the two kinds of material. A *t* test for independent samples was done to determine the difference in ability as exhibited by both groups, first for the structured lists and then for the unstructured lists. The good readers learned the structured lists more rapidly than the poor readers ($t = 4.89$, $p < .001$). In the case of the unstructured material, both groups learned the lists with equal degree of difficulty ($t = 1.53$, $p = \mathit{ns}$).

TABLE 3

MEAN NUMBER OF TRIALS TO CRITERION

Reader	Structured lists	Unstructured lists
Good		
M	10.02	17.83
SD	6.28	5.06
Poor		
M	17.00	19.97
SD	4.61	3.67

Statistical Procedures to Control for the Effect of Intelligence

Though every effort was made to equate the good and poor readers on a measure of intelligence, the IQ scores of the good readers were significantly higher. Since the Henmon-Nelson test is a group intelligence measure which requires the subject to read the items, and research findings have indicated the wide discrepancy in IQ points between group tests and individual tests administered to disabled readers (Schiffman, 1962), it is highly probable that the good readers enjoyed a definite advantage. Nevertheless, it was decided to investigate the extent to which statistically significant differences in performance on the task between good and poor readers could be attributed to differences in IQ test scores. An analysis of covariance was computed to partial out the effects of intelligence on the following measures: (*a*) the Gates reading scores, (*b*) difference scores for good and poor readers—mean trials to criterion for structured lists compared to the mean for the unstructured lists, and (*c*) mean trials for structured material and unstructured material compared separately across the two groups.

The good and poor readers remain significantly differentiated as to their reading ability, $F = 93.35$, $p < .001$. The difference score is a measure which essentially combines two earlier measures—a comparison of the effect of syntactic structure on the performance of both good and poor readers. Earlier, a correlated t test was used for each group. With intelligence partialed out, the facilitative effect of syntactic structure is greater for good readers than for poor readers, $F = 7.16$, $p < .05$. Again, the difference between a child's performance on structured material and unstructured material varies directly and reliably with his reading ability.

The difference between good and poor readers on the structured lists is significant, $F = 18.05$, $p < .001$, whereas performance on the unstructured lists does not differentiate the good readers from the poor readers ($F = .02$) as was found in the earlier analysis.

DISCUSSION

The primary focus of this investigation is the relation of grammatical structures to reading ability. The poor readers in the present study were in fact almost average readers—they were Grade 4 students reading at a Grade 4.04 level. It would have served the experimental paradigm to select children with definite reading difficulties. However, this involved the high risk of including some children with obvious or subtle sensory problems, and many children with emotional or attitudinal problems related to learning. It would also have resulted in even greater IQ score discrepancies between the good and poor readers. One could argue that the results of the present study are the more impressive and relevant to reading ability in general because differences in the use of syntactic cues were evident in a comparison of good readers and less good readers, without resorting to extremes in the selection of subjects. It may be, however, that any generalization of these results to children with severe reading disabilities is tenuous.

The fact that the good and less good readers in the present sample did equally well on the paired-associate learning task suggests that the poorer readers can learn to associate simple familiar words with the ease of good readers. Studies of paired-associate learning with normal and retarded readers present contradictory findings. Otto (1961) found that good, average, and poor readers, in that order, need increasingly more trials to learn a list of paired associates. His paradigm combined both visual and auditory presentation, for example, a picture of a diamond and the nonsense word "fep", thus complicating the association. Budoff and Quindlan (1964), using only auditory presentation of meaningful simple four-letter words, found that their disabled readers did better in a paired-associate learning situation than the normal readers. The present findings of no difference in paired-associate learning between the two reading groups is consistent with the report of Budoff and Quindlan (1964). It is further supported by the finding that in learning the unstructured lists of words the good readers do no better than the less good readers.

The difference between the two groups of readers is their ability to make use of the information inherent in the grammatical structure of a sentence. If we use Braine's (1963) concept of "contexual generalization" to explain the acquisition of word order, then we require the mediating property of generalization to be extrinsic, namely, temporal location in an utterance. There exist certain elements that recur in an infinite number of sentences—these elements reduce the complexity of defining position. The child learns to recognize these markers and they begin to function as reference points. Learning and generalization are assisted because words which appear in a given position are tagged by the marker specific to that position; tagged either by the function word or bound morpheme ending. The speculation in the present report is that the poorer readers have not generalized the grammatical markers. The good readers, having experienced an "ed" ending in a particular location in many sentences, will be able to perceive the syntactic structure of the "ed" marker in a novel situation where the stem to which it is attached is unfamiliar, that is, without meaning. The subjects in this study had to get information from the markers or learn the lists by rote association. It would appear that both good and poor readers approached the unstructured material in this rote manner but parted ways when the opportunity to use grammatical cues was available.

This study raises a number of questions that merit attention. Eisenberg (1966, p. 7) indicates that the greatest incidence of reading retardation occurs among the lower socioeconomic classes. It would be interesting to compare lower- and middle-class poor readers on their ability to use syntactic cues. The results of such studies might have implications for compensatory programs.

Another general question raised by the present study relates to the assumptions underlying procedures used for language and reading remediation. Difficulties in following verbal directions and dropped endings in reading are common problems among children with learning disabilities which are usually explained in perceptual rather than cognitive terms. It would be worthwhile trying to develop and test remedial procedures based on the cognitive-linguistic model which proved to be useful in the present study.

REFERENCES

BRAINE, M. On learning the grammatical order of words. *Psychological Review*, 1963, **70**, 323–348.

BUDDOFF, M., & QUINDLAN, D. Reading progress as related to efficiency of visual and oral learning in the primary grades. *Journal of Educational Psychology*, 1964, **55**, 247–252.

EISENBERG, L. The epidemiology of reading retardation and a program for preventive intervention. In J. Money (Ed.), *The disabled reader.* Baltimore: Johns Hopkins, 1966.

EPSTEIN, W. The influence of syntactic structure on learning. *American Journal of Psychology*, 1961, **74**, 80–85.

EPSTEIN, W. A further study of the effect of syntactical structure on learning. *American Journal of Psychology*, 1962, **75**, 121–126.

FRIES, C. C. *Linguistics and reading*. New York: Holt, Rinehart & Winston, 1962.

FORSTER, K. I. The effect of syntactic structure on nonordered recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 292–297.

GATES, A. I. *Gates reading survey*. New York: Columbia University, Teachers College, 1958.

HENMON, V. A. C., & NELSON, J. N. *Manual for The Henmon-Nelson Tests of Mental Ability*. Boston: Houghton-Mifflin, 1958.

LASHLEY, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. (The Hixon Symposium) New York: Wiley, 1951.

MUEHL, S., & KREMENAK, S. Ability to match information within and between auditory and visual sense modalities and subsequent reading achievement. *Journal of Educational Psychology*, 1966, **57**, 230–238.

ORTON, S. T. *Reading, writing, and speech problems in children*. New York: Norton, 1937.

OTTO, W. The acquisition and retention of paired associates by good and poor readers. *Journal of Educational Psychology*, 1961, **52**, 241–248.

SCHIFFMAN, G. Dyslexia as an educational phenomenon. In J. Money (Ed.), *Reading disability progress and research needs in dyslexia*. Baltimore: Johns Hopkins, 1962.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Columbia University, Teacher's College, 1944.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. New York: J. B. Lippincott, 1960.

(Received November 7, 1969)

Journal of Educational Psychology
1971, Vol. 62, No. 1, 31–38

INTERACTIVE RELATIONSHIPS AMONG LEARNER CHARACTERISTICS, TYPES OF LEARNING, INSTRUCTIONAL METHODS, AND SUBJECT MATTER VARIABLES[1]

G. KASTEN TALLMADGE[2] AND JAMES W. SHEARER
American Institutes for Research in the Behavioral Sciences, Palo Alto, California

This study is the most recent of a series of investigations aimed at determining whether learning can be enhanced by employing instructional treatments differing in design and use as a function of learner characteristics. Six experimental courses were developed reflecting different combinations of inductive and deductive teaching methods, rote and understanding types of learning, and two different subject matters. Achievement test scores following instruction were correlated with the 39 aptitude, interest, and achievement measures obtained on the 353 Navy enlisted men who served as experimental subjects. A statistically significant interactive relationship was found between learner anxiety and method of instruction which was consistent across the two subject matters and the two types of learning.

This study was the fourth in a series of research projects aimed at determining whether learning might be enhanced by employing instructional methods which differed in design and use as a function of learner characteristics. The first study of the series (Tallmadge, 1968) contrasted two experimental versions of a 1 week course in Maneuvering Board with the standard version of the course in the Radarman Class A School at Treasure Island, California. It produced no significant interactions between instructional method and learner characteristic variables, although some indirect evidence was produced suggesting that the nature of the material to be learned was critical and needed to be examined as a separate, independent experimental variable.

This hypothesis was the basis for the second study (Tallmadge & Shearer, 1969) which was designed to examine higher order interactions among learner characteristics, instructional methods, and "subject matters." Inductive and deductive experimental courses were developed for a logico-mathematical subject matter (Transportation Technique) and a visual form discrimination topic (Aircraft Recognition). For both courses interactions occurred between noncognitive learner characteristics and methods of instruction. It was found, however, that the direction of the relationships which existed between learner characteristics and methods of instruction for the Transportation Technique courses were reversed for the Aircraft Recognition courses. In other words, those subjects who did well in the inductive version of the Transportation Technique course had the same characteristics as subjects who did well in the deductive version of the Aircraft Recognition course. Similarly, subjects who performed well under the inductive version of the Aircraft Recognition course had the same characteristics as those who performed well under the deductive version of the Transportation Technique course.

Later, as part of an attempt to build a model relating types of subject matter, types of learning, and types of teaching to achievement differences, Tallmadge and Shearer (1968) noted and described how type of subject matter and type of learning had been confounded in the Tallmadge and Shearer (1969) study. For the "meaningful rules" subject matter (Transportation Technique), the inductive course had taught concepts and principles in addition to problem-solving procedures. The deductive course, on the other hand, was limited to coverage of the problem-solving procedures. Thus it was not possible to determine whether the observed interaction for the Transportation Technique topic was between learner characteristics and instructional methods or between learner characteristics and types of learning.

---

[1] This report is based on research conducted under Contract N61339-68-C-0271 with the Naval Training Device Center, Orlando, Florida, and monitored by G. S. Micheli. A more detailed description of this research is available as Technical Report NAVTRADEVCEN 68-C-0271-1. The opinions or conclusions expressed herein are those of the authors and should not be construed as necessarily reflecting the views or endorsement of the Navy Department or the Naval Training Device Center. The authors gratefully acknowledge the assistance of A. O. H. Roberts, O. A. t'Hooft, and B. M. Lauritz for their assistance in various phases of the work.

[2] Requests for reprints should be sent to G. Kasten Tallmadge, Instructional Methods Program, American Institutes for Research, P. O. Box 1113, Palo Alto, California 94302.

The primary research concern of the present study was resolving the question of whether type of teaching (inductive versus deductive) or type of learning (understanding versus rote) was responsible for the interaction effects observed by Tallmadge and Shearer (1969). A second research concern was partial verification of the hypothesis that the "meaningful rules–arbitrary rules" dichotomy was responsible for the observed reversal of Learner Characteristic × Instructional Method interactions which occurred as a function of training topic differences. Finally, it was hoped that the study would both verify inferences made about relevant learner characteristic variables and sharpen their measurement. The study employed the basic model shown in Figure 1.

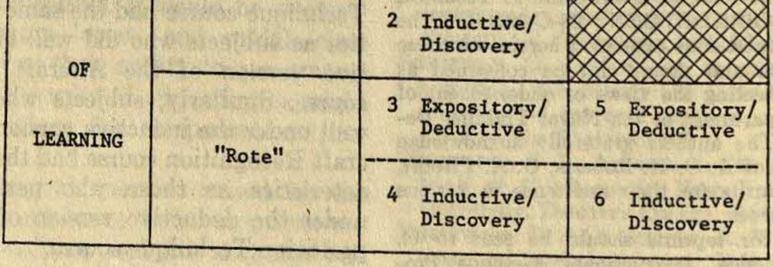
Method

The basic experimental design model shown in Figure 1 was first described by Tallmadge and Shearer in 1968. As shown in Figure 1, the design permits independent examination of two different subject matters, two different types of learning, and two different types of teaching, as well as interactions among these variables. A fourth dimension, learner characteristics, is not shown in the model but is, of course, an essential ingredient of any research of this type.

Because there can be no such thing as "understanding learning" of arbitrary rules subject matter, the Figure 1 model does not represent a complete factorial design. It may, however, be regarded as two overlapping complete factorial designs. For the meaningful rules subject matter, types of learning and types of teaching constitute one 2 × 2 complete factorial design. Within the rote type of learning, subject matters and types of teaching constitute a similar design.

Several criteria were established for selection of the two subject matter topics to be developed and taught experimentally. Minimum requirements were that one course should cover a meaningful rules topic and the other an arbitrary rules topic and that each should be as homogeneous as possible with respect to types of skills and knowledges required (e.g., factorially simple). It was also necessary that both subject matters be amenable to both an inductive or discovery instructional approach and an expository-deductive approach.

Four additional criteria were generated as a result of practical considerations: (*a*) the selected



		SUBJECT MATTERS	
		Meaningful Rules	Arbitrary Rules
TYPES OF LEARNING	Understanding	1 Expository/ Deductive	▨
		2 Inductive/ Discovery	▨
	"Rote"	3 Expository/ Deductive	5 Expository/ Deductive
		4 Inductive/ Discovery	6 Inductive/ Discovery

Fig. 1. Model showing relationships among instructional methods, subject matters, and types of learning.

courses had to be amenable to group administration, (*b*) large and/or expensive equipment requirements had to be avoided, (*c*) prerequisite skills had to be minimal to avoid extensive pretraining or the need to use highly selected subject populations, and (d) the topics had to be "new" to the subjects. After considerable investigation it was concluded that Celestial Navigation and Aircraft Recognition would satisfy all criteria relevant to the selection of material for experimental training.

Based on previously specified sets of behavioral objectives, four experimental Celestial Navigation courses were developed for use in the study: Inductive-Discovery-Understanding, Expository-Deductive-Understanding, Inductive-Discovery-Rote, and Expository-Deductive-Rote. The courses were designed to teach subjects how to set up and solve simple navigation problems using procedures adapted specifically for the study. Detailed descriptions of these courses are presented by Tallmadge and Shearer (1970). The inductive and deductive Aircraft Recognition courses were essentially identical to those employed in the Tallmadge and Shearer (1969) study.

Criterion tests were developed for the two subject matter areas directly from the statements of behavioral objectives and prior to course development. A single criterion test was used for all courses within each training topic.

The subjects for the study were 353 Navy enlisted men awaiting assignment to Basic Electricity and Electronics School in San Diego. A battery of psychological tests was administered to groups of approximately 30 subjects during the first day of the 2-day experimental period. The test battery included the Kuder Vocational Preference Record (Form BB), Gordon Personal Profile, Eysenck Personality Inventory (Form A), Sixteen Personality Factor Questionnaire (16 PF, Form B), Set Toward Education Scale (Capretta, Jones, Siegel, & Siegel, 1963), Gestalt Completion Test, and the Hidden Figures Test. The selection of these instruments was based on evidence obtained from earlier studies in this series and on a review of the research literature. In addition, scores on the Navy Basic Battery General Classification, Arithmetic, Mechanical, and Clerical tests were obtained for each subject. These tests brought the number of obtained individual difference measures to a total of 39.

The six experimental courses were taught at Naval Training School facilities with project personnel serving as instructors. Each course was taught, in a random sequence, to two (or in one case, where subject attrition was high, to three) groups of subjects. The experimental classroom sessions consumed approximately 1 full day (the Celestial Navigation courses required slightly longer than the Aircraft Recognition courses).

Criterion tests were administered to all subjects by project staff members immediately following completion of instruction. The Celestial Navigation criterion examination required 1 hour and 20 minutes. It consisted of some 17 problems, the majority of which involved several intermediate steps. Partial credit was given for partial solutions. This scoring system provided a possible range of scores from 0 to 53.

The Aircraft Recognition criterion examination required approximately 30 minutes. It consisted of 56 aerial photographs depicting 21 different aircraft which the subjects were instructed to identify.

Raw scores on the two criterion tests were standardized ($\bar{X} = 50, SD = 10$) prior to data analysis. This standardization served the purpose of removing all differences between the two subject matter areas. The original differences in raw scores could only be attributed to the lengths and difficulty levels of the criterion tests—not to any inherent differences between the subject matters—and could be seriously misinterpreted if not removed in this manner.

RESULTS

Correlation coefficients were calculated, using the BMD-03D computer program, between predictor and criterion test scores for all 39 predictors in each of the six treatment conditions.

With but few exceptions the correlations were of approximately the anticipated magnitude. One notable exception was the correlation ($r = .60$) between the Set Toward Education Scale and achievement under the Celestial Navigation Inductive Rote treatment condition. The $p = .95$ confidence interval for this correlation was found to be $.41 < p < .74$ indicating a substantial relationship. The corresponding correlations under other treatment conditions were, for the most part, not significantly different from zero.

The Kuder Computational and Scientific scales also showed surprisingly high correlations with achievement ($r = .50$ and $r = .52$, respectively) for the Celestial Navigation Deductive Understanding treatment

TABLE 1

CORRELATIONS SHOWING INTERACTIONS BETWEEN MEASURES OF ADJUSTMENT AND INSTRUCTIONAL METHODS

Measure	Inductive	Deductive
Eysenck Neuroticism	.11	−.17
16 PF Emotionally Stable	−.13	.16
16 PF Apprehensive	.12	−.16
Multiple	.14	−.20

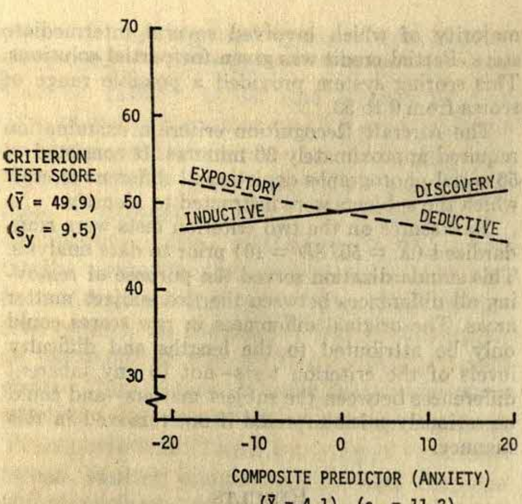


Fig. 2. Interaction between anxiety and instructional method.

condition. Neither of these coefficients was significantly different from the corresponding correlations for the other three Celestial Navigation treatments, however.

Of primary interest were differences between correlations existing as a function of treatment conditions. No consistent patterns of this type were apparent through visual inspection. Similarly, tests of differences of pairs of correlations drawn from different treatment groups did not serve to identify meaningful patterns although a number of statistically significant differences were found.

To increase the power of statistical tests for subsequent analyses, treatment groups were merged to form combinations appropriate for separate examination of each treatment dimension. Comparisons between understanding and rote types of learning and between the Celestial Navigation and Aircraft Recognition subject matters did not produce any correlational differences indicative of learning style effects. When the three inductive groups were pooled and compared against the pooled deductive groups, however, a consistent and apparently meaningful interactive relationship between learner characteristics and instructional methods was discovered.

Three of the scales which showed significant interactive relationships with the instructional method variable (Eysenck Neuroticism, 16 PF Emotionally Stable, and 16 PF Apprehensive) were essentially measures of adjustment. The fact that these scales interacted in the same manner with instructional methods lent considerable credibility to the existence of an interaction between student anxiety and method of instruction. Table 1 presents the relevant correlations as well as multiple correlations for the inductive and deductive groups which were computed by assigning weights of $+1$ to the Eysenck Neuroticism and 16 PF Apprehensive scales, and a weight of -1 to the 16 PF Emotionally Stable scale.

The regression lines corresponding to the multiple *r*s of Table 1 are plotted in Figure 2. This figure illustrates that the regression lines cross near the predictor score mean. The slopes of the regression lines were found to differ significantly $(p < .05)$ as did the multiple *r*s. To further investigate this interactive relationship, the subjects were subsequently sorted into high and low groups depending upon whether their composite predictor score fell above or below the point where the regression lines crossed $(X = -.36)$. Finally, mean achievement test scores were computed for those above the crossover point who received an inductive treatment, those above the crossover point who received a deductive treatment, those below the crossover point who received an inductive treatment, and those below the crossover point who received a deductive treatment. These mean scores (together with the associated standard deviations) are presented in Table 2.

Finally, an analysis of variance was performed on the criterion test scores for the

TABLE 2

Criterion Test Score Means and Standard Deviations for High- and Low-Anxiety Groups Taught Inductively and Deductively

Anxiety	Inductive	Deductive
High		
$\bar{X}$	50.64	48.63
SD	9.85	8.94
Low		
$\bar{X}$	48.27	52.14
SD	8.91	10.19

four groups using the unweighted means technique (Winer, 1962). The results of this analysis are presented in Table 3. As can be seen, neither the learner characteristic nor the instructional method main effect was statistically significant while their interaction was significant.

Differences between pairs of means shown in Table 2 were tested using Student's t. The deductive instructional method was found to be significantly superior ($p < .05$) for the low-anxiety group. The instructional treatments did not produce significant achievement differences ($p > .10$) for the high-anxiety group. Similarly, the low-anxiety group performed significantly better than the high-anxiety group ($p < .05$) under the deductive treatment while the difference between groups was not statistically significant ($p > .10$) under inductive instructional conditions.

The experimental Aircraft Recognition courses used in this research were essentially identical to the Aircraft Recognition courses used in an earlier study in this series (Tallmadge & Shearer, 1969). The Aircraft Recognition portion of this study, then, constituted one of few replications of a study in which significant Learner Characteristic × Treatment interactions had been found. The earlier study found that the inductive and deductive treatments produced significantly different correlations with the criterion for three Kuder Vocational Interest scales (Computational, Scientific, and Social Service), and two Gordon Personal Profile scales (Ascendancy and Sociability). Not one of the differences between these same pairs of correlations was statistically significant in the present study although, with the exception of Gordon Ascendancy, the direction of the differences was the same in both studies.

TABLE 3

UNWEIGHTED MEANS ANALYSIS OF VARIANCE

Source	*df*	*MS*	*F*	*p*
Instructional methods	1	79.5	<1	
Learner characteristics	1	27.0	<1	
Interaction	1	710.0	8.0	<.01
Error	349	88.7		

TABLE 4

MEAN ACHIEVEMENT SCORES

Group	Aircraft Recognition Earlier study		Aircraft Recognition Present study	
	Inductive	Deductive	Inductive	Deductive
High	46.68	53.61	50.16	50.34
Low	52.19	48.04	49.17	49.15

In the earlier study, a composite measure consisting of the Kuder Scientific Interest scale (positively weighted), the Gordon Ascendancy scale (negatively weighted), and the Kuder Musical Interest scale (negatively weighted) was developed to maximize between-treatment-group differences. The subjects scoring above the median on this composite measure showed significantly greater (Aircraft Recognition) achievement under deductive than inductive instruction ($p < .001$). The subjects scoring below the median performed better under inductive instruction although not significantly so.

The same composite score was calculated for subjects taught Aircraft Recognition in the present study. Using these data, no between-treatment differences were found for subjects scoring either above or below the median. Table 4 presents a comparison of data from the two studies.

Predictor and criterion test score means and variances were found not to differ significantly between replications nor were comparisons significantly affected by using the composite score median from the earlier study to sort present study subjects into high and low groups. The most plausible explanation as to why the two replications produced different results was that different instructors taught the two courses in the earlier study while a single instructor taught both courses in this study. Consistent with findings reported by Heil and Washburn (1961), instructor rather than instructional method differences may have produced the observed interaction.

While the primary concern of this study was with interactive relationships involving learner characteristic variables, it did not seem appropriate to ignore completely any overall effectiveness differences which might have existed between treatment conditions.

TABLE 5

MEAN ACHIEVEMENT SCORES AND STANDARD DEVIATIONS FOR THE SIX TREATMENT CONDITIONS

Condition	Inductive	Deductive
Celestial Navigation Understanding		
$\bar{X}$	50.6	49.0
SD	8.8	9.9
Celestial Navigation Rote		
$\bar{X}$	49.2	51.1
SD	10.1	8.8
Aircraft Recognition		
$\bar{X}$	49.6	49.6
SD	9.7	9.7

For this reason, mean achievement scores and associated standard deviations were calculated and are presented in Table 5.

The F_{MAX} test was used to test for homogeneity of variance. No significant difference was found ($F_{\text{MAX}} = 1.21$). Differences between all pairs of means were tested using *t* tests (a conservative procedure when attempting to minimize Type 1 error). No statistically significant differences were found.

While it should be pointed out that between-subject-matter differences were statistically eliminated by standardizing scores within subject matters, results of the above described tests led to the conclusion that within-subject-matter treatment differences produced no significant achievement differences.

DISCUSSION

The purpose of this study was to demonstrate the practical significance of Learner Characteristic × Instructional Treatment interactions by verifying the experimental findings reported in a previous study (Tallmadge & Shearer, 1969) and by attempting to increase the magnitude of the observed relationships through an improved experimental design and sharpened measurement techniques. It is readily apparent from the results described above that this goal was excessively optimistic and that far more basic knowledge is required before learning style research is likely to have a significant impact on any real-world training programs.

At the time the study described here was undertaken, other investigators shared with the authors of this article a general feeling of enthusiasm for the practical implications of research in this area. Numerous studies were appearing which reported significant Learner Characteristic × Treatment interactions, and it seemed that soon some order would emerge from the apparently unrelated and frequently discrepant findings.

Two recent reports (Bracht 1969; Cronbach & Snow, 1969) which have reviewed, critiqued, and summarized the published literature revealed, however, a more cautious attitude toward the existence of meaningful interactions between learner characteristic and instructional method variables than was prevalent only a year ago. Bracht's literature review summarized some 90 studies of which only 5 produced interactions which met his criteria for meaningfulness.

Two of the five meaningful interactions reported by Bracht involved personality characteristics. These findings led him to suggest that the personality domain was deserving of further investigation. Cronbach and Snow (1969) were less enthusiastic although Cronbach's (1966) earlier position was that "the interacting variables may have more to do with personality than with ability [p. 90]." In their 1969 report, these authors decried learning style studies involving personality variables describing them as "disappointing and unencouraging [p. 173]." At the same time, however, they described a number of studies reporting significant personality characteristic interactions. In fact, on page 159 of their report, they indicated that they had found repeated hints in the literature of an interaction between anxiety and instructional treatment variables. This conclusion was consistent with the findings of the present study as well as the Tallmadge and Shearer (1969) research.

Despite the existence of some apparent consistency in the research literature, an overview would certainly uncover more negative than positive findings and more inconsistencies than consistencies. If different learning styles exist, they are indeed elusive —a conclusion amply demonstrated by the research of Bunderson (1969). After revising an instructional treatment to simplify it and reduce the time consumed by its administration, he found that Aptitude × Treatment interactions obtained with the original instruction were *reversed* for the revised version. A study by Burton and Goldbeck (1962) offers further evidence supporting the unpleasant possibility that reducing the difficulty of the learning task may reverse aptitude-treatment relationships.

As the proliferation of research results produced an increasing number of closely comparable studies, the expected emergence of consistent patterns of relationships simply did not occur. Rather, it appeared that relatively minor differences between treatments could cause a complete reversal of interactive relationships. These findings, coupled with the large number of studies reporting negative results or results inconsistent with expectations, have caused the prevailing mood to change from optimism to pessimism. Bunderson (1969), for example, has decried the "robustness" of Aptitude × Treatment interactions suggesting that other approaches are more likely to improve instructional effectiveness. Certainly the results of the present study are, for the most part, consistent with this recently expressed viewpoint.

Still, it does not appear that learning style research should be abandoned. The relationship found in the present study between anxiety and (inductive versus deductive) teaching methods, while not impressively large, was sufficiently robust to stand up for two different types of learning, two different subject matters, and probably at least two difficulty levels.

The single positive finding of the present study which the authors feel is reliable and which should stand up to replication was the above-mentioned interactive relationship between learner anxiety and method of instruction. While many "statistically significant differences" were found between pairs of predictor-criterion correlation and regression coefficients as a function of treatment conditions, the majority of these differences were considered to be unreliable since they represented only a small fraction of the total number of pairs tested, and since they formed no clear-cut pattern.

The significant negative findings of the present study were that neither subject matter nor type of learning appeared to be related to Learner Characteristics × Instructional Methods interactions. These findings apparently negated the potential utility of further investigation of these variables. While the authors would not accept this position in the absence of replication or other research specifically concerned with subject matter variables, they would concur that investigations of learner-centered variables appear more promising at the present time.

Nearly all research in the learning style area has employed individual difference measures developed for other, usually specific, applications. While those studies which tested hypotheses about aptitude variables have been well served by such instruments, studies concerned with noncognitive learner characteristics, especially personality variables, have been less fortunate. For the most part, studies of the latter type have been exploratory rather than hypothesis-oriented. They have produced statistically significant findings with at least the same frequency as aptitude-related studies but have been difficult to interpret, and especially to compare, because of possible nonlinearity of regression, differences in "class medians" between studies, and other contaminating factors.

It is the authors' belief that the most significant future developments in the learning style area will involve noncognitive learner characteristics. They also believe that specially developed measuring instruments based either on item analyses of existing tests or on specifically formulated hypotheses will be most likely to produce meaningful results.

Beswick (1969) has recently pointed out certain parallels between the learning style research reported by Tallmadge and Shearer (1969) and some of his own research on curiosity (Beswick, 1964, 1968). He has suggested that possible interactions between intrinsic motivation (curiosity) and the conditions of instruction may be attributable to "conceptual conflict" with both subject matter and instructional method acting as moderator variables.

According to Beswick's theory, subjects high on the curiosity continuum will be positively motivated by instructional treatment conditions high in "conceptual conflict"—conditions which might be described as intellectually challenging—and negatively motivated by low "conceptual conflict." The reverse pattern would be predicted for subjects low in curiosity. Clearly, relationships between conceptual conflict and type of learning, type of subject matter, and method of instruction would not necessarily be unidirectional, and the latter variables could interact in many ways to affect the conceptual conflict of an instructional treatment.

Beswick's hypothesis regarding the motivational effects of conceptual conflict on subjects high and low in curiosity would readily account for the results observed by Bunderson (1969) and by Burton and Goldbeck (1962) as a function of treatment difficulty differences. Further investigation might reveal other apparently conflicting results which could be explained by his conception of curiosity.

Whether or not Beswick's complex formulation of the learning style issue proves valid, it is the kind of theoretical approach which the authors believe holds promise of leading eventually to a resolution of the many still unanswered questions. They feel it is now clear that simple solutions will not be found and that, unless research is designed to test specific hypotheses, much effort will be wasted in speculation over the many seeming inconsistencies in the research literature.

REFERENCES

Beswick, D. G. Theory and measurement of human curiosity. Unpublished doctoral dissertation, Harvard University, 1964.

Beswick, D. G. Correlates of individual differences in curiosity. Paper presented at the annual conference of the Australian Psychological Society, Brisbane, August 1968.

Beswick, D. G. Notes on learning styles, instructional methods, and learning experiences. Unpublished manuscript, Australian National University, 1969.

Bracht, G. H. The relationship of treatment tasks, personological variables, and dependent variables to aptitude-treatment interactions. Unpublished doctoral dissertation, University of Colorado, 1969.

Bunderson, C. V. Ability by treatment interactions in designing instruction for a hierarchial learning task. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, February 1969.

Burton, B. B., & Goldbeck, R. A. *The effect of response characteristics in multiple choice alternatives on learning during programed instruction.* (AIR-C28-2/62-TR) San Mateo, Calif.: American Institutes for Research February 1962.

Capretta, P. J., Jones, R. L., Siegel, L., & Siegel, L. C. Some noncognitive characteristics of honors program candidates. *Journal of Educational Psychology,* 1963, **54,** 268–276.

Cronbach, L. J. The logic of experiments on discovery. In L. S. Shulman & E. R. Keislar (Eds.), *Learning by discovery: A critical appraisal.* Chicago: Rand McNally, 1966.

Cronbach, L. J., & Snow, R. E. Individual differences in learning ability as a function of instructional variables. (United States Office of Education Final Report, Contract OEC 4-6-061269-1217) Stanford, Calif.: Stanford University, 1969.

Heil, L. M., & Washburne, C. Characteristics of teachers related to children's progress. *Journal of Teacher Education,* 1961, **12,** 401–406.

Tallmadge, G. K. Relationships between training methods and learner characteristics. *Journal of Educational Psychology,* 1968, **59,** 32–36.

Tallmadge, G. K., & Shearer, J. W. *Study of training equipment and individual differences: The effects of subject matter variables. Supplementary Report.* (Technical Report No. NAVTRADEVCEN 67-C-0114-2) Orlando, Fla.: Naval Training Device Center, August 1968.

Tallmadge, G. K., & Shearer, J. W. Relationships among learning styles, instructional methods, and the nature of learning experiences. *Journal of Educational Psychology,* 1969, **60,** 222–230.

Tallmadge, G. K., & Shearer, J. W. *Study of training equipment and individual differences: Research on interactive relationships among learner characteristics, types of learning, instructional methods, and subject matter variables.* (Technical Report No. NAVTRADEVCEN 68-C-0271-1) Orlando, Fla.: Naval Training Device Center, June 1970.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received November 16, 1969)

Journal of Educational Psychology
1971, Vol. 62, No. 1, 39–44

RETROACTIVE INTERFERENCE IN READING PROSE MATERIALS[1]

JAMES H. CROUSE[2]
University of Delaware

Three experiments investigated retroactive interference phenomena with prose-learning tasks. Experiment I dealt with the effects of interpassage similarity on retroactive interference. Interpassage similarity was defined in terms of experimenter-generated questions which the successive passages answered. The passages were constructed so that, in successive passages, the experimenter-generated questions were different, similar, or identical, but always had different answers. No retroactive interference (relative to a control) was found in the Different condition, but approximately 40% retroactive interference was obtained in Similar and Identical conditions. Experiment II supported the conclusion that the retroactive interference was not a result of failures of passage differentiation, and Experiment III suggested that it was not a result of output phenomena.

A basic procedure for studying reading is the typical single-trial prose-learning task. This task involves three phases; a phase in which the subject is instructed to read a passage, an input phase in which the passage is presented, and an output phase in which questions are asked. In a recent study (Crouse, 1970), the subjects in an experimental condition read one passage during original learning, then another passage during interpolated learning, and then recalled the original learning passage. Recall of the original learning passage in this experimental condition was lower than in a control condition which read the original learning passage, then had a filler task for the time required for interpolated learning, and then recalled the original learning passage. This decrement, known as retroactive inhibition, duplicated for prose-learning tasks the common finding for a variety of other learning tasks; for example, free-recall, paired-associate, and serial learning tasks (cf. Keppel, 1968). The same general retroactive inhibition paradigm was followed in the experiments reported in the present study with the addition that two interpolated learning passages rather than one were used, a procedure which generally increases the amount of retroactive inhibition obtained (cf. Postman, 1965).

EXPERIMENT I

The purpose of Experiment I was to determine whether the amount of retroactive inhibition in prose-learning tasks would be affected by the similarity of materials learned in the successive tasks; that is, interpassage similarity. Retroactive inhibition generally increases with the similarity of materials learned in successive tasks for free-recall, paired-associate, and serial learning (cf. Keppel, 1968). The approach taken here was to develop fairly short passages, and for each passage the experimenter generated a relatively large set of questions that the passage answered; thus, the important feature of each passage was the set of experimenter-generated questions answered by the passage. Three conditions of interpassage similarity were defined. In the Different condition passages were constructed so that, in successive passages, the experimenter-generated questions were different and had different answers, in the Similar condition the experimenter-generated questions were similar and had different answers, and in the Identical condition the experimenter-generated questions were identical and had different answers. It was expected the retroactive inhibition would increase from the Different condition to the Similar condition to the Identical condition.

---

[1] This research was supported in part by United States Public Health Service Research Grant MH16909-02 from the National Institute of Mental Health.

[2] Requests for reprints should be sent to James H. Crouse, College of Education, University of Delaware, Newark, Delaware 19711.

Method

Design. There were four conditions in which the subjects learned the same passage during original learning. Three were experimental conditions (Different, Similar, and Identical) in which the original learning passage was learned, then an interpolated learning passage was learned (Interpolated Learning 1) then a second interpolated learning passage was learned (Interpolated Learning 2), and then original learning recall was taken. The fourth was a control condition for forgetting in which the original learning passage was learned, then a filler task was performed for the time required for interpolated learning, and then original learning recall was taken. Twenty subjects, undergraduate education students at the University of Delaware, were assigned to each of the four conditions.

Passages. The original learning passage was a 212-word passage about a hypothetical person, John Payton, in which the first two sentences were: "Payton was born in Hampstead at the end of October, 1795. When he was only eight years of age, his father, who kept a livery stable, was killed by a fall from a horse." A set of 22 questions was generated which asked, by inspection, a large number of questions that could be answered by the passage. Each question could be answered unambiguously by one or more words from the passage. The three experimental conditions were distinguished only by the interpolated passages used. The three experimental conditions were:

1. *Identical condition.* The Interpolated Learning 1 passage was a 205-word passage constructed by making one change in the original learning passage; the segments of the original learning passage which answered the 22 questions were replaced with new answers. The first two sentences were: "Payton was born in Liverpool at the end of October, 1810. When he was only five years of age, his father, who was a servant, was killed by a robber." As a result of these changes, a set of 22 questions was generated for the Interpolated Learning 1 passage. These questions were identical to those for the original learning passage, but the answers were different. For example, the question "Payton was born in what city?" was answered by "Hampstead" in the original learning condition and "Liverpool" in Interpolated Learning 1. The Interpolated Learning 2 passage was a 206-word passage constructed from the Interpolated Learning 1 passage in the same way the Interpolated Learning 1 passage was constructed from the original learning passage; that is, the segments of the Interpolated Learning 1 passage which answered the 22 questions were replaced with new answers. As a result of these changes, a set of 22 questions was generated for the Interpolated Learning 2 passage which were identical to those for the original learning and Interpolated Learning 1 passages, but the answers were different.

2. *Similar condition.* The Interpolated Learning 1 passage was a 205-word passage constructed from the original learning passage by making two changes; the segments of the original learning passage which answered the 22 questions were replaced with the answers of the Interpolated Learning 1 passage in the Identical condition, and the name of the hypothetical person was changed from Payton to Fowler. The first two sentences were: "Fowler was born in Liverpool at the end of October, 1810. When he was only five years of age, his father, who was a servant, was killed by a robber." As a result of these changes, a set of 22 questions was generated which were similar to those of the original learning, but the answers were different and the same as the Interpolated Learning 1 answers in the Identical condition. For example, the question "Payton was born in what city?" was answered by "Hampstead" in the original learning task, and the question "Fowler was born in what city?" was answered by "Liverpool" in the Interpolated Learning 1 task. The Interpolated Learning 2 passage was a 206-word passage constructed from the Interpolated Learning 1 passage in the same way the Interpolated Learning 1 passage was constructed from the original learning passage; that is, the segments of the Interpolated Learning 1 passage that answered the 22 questions were replaced with the answers of the Interpolated Learning 2 passage in the Identical condition, and the name of the hypothetical person was changed from Fowler to Hughes. As a result of these changes, a set of 22 questions was generated which were similar to those of the original learning and Interpolated Learning 1 passages, but the answers were different and the same as the Interpolated Learning 2 answers in the Identical condition.

3. *Different condition.* The Interpoloted Learning 1 passage was a 204-word passage about a hypothetical island, Karisoon. The first two sentences were: "Karisoon (population 29,723) is an island whose main attraction is recreation. Even though it is a small island its many inlets give it a tidal coast of 1,057 miles." The set of 22 questions generated for this passage was different from the original learning set and had different answers. The Interpolated Learning 2 passage was a 214-word passage about a hypothetical library, King Library. The first two sentences were: "King Library is particularly well situated being in the upper section of beautiful Northwark Park. Each year the library is host to several art exhibits, the largest being the Kenwood exhibit." The set of 22 questions generated for this passage was different from the original learning and Interpolated Learning 1 sets and had different answers.

Procedure. The subjects were run in small groups of three to seven which were randomly assigned to the four conditions. The materials were combined into booklets and the learning procedure for each passage was as follows. During input the subject was presented a passage on a single page for 2 minutes. Immediately following input he was presented with questions on a single page to which he printed his answers. In order to provide the greatest opportunity for output, the entire set of 22 questions for a passage was presented. The order of the questions on the page was "scrambled"; that is, it did not follow the sequence of the passage. For each condition, the same scrambled order was used for original learning output, another scrambled order was used for Interpolated Learning 1 output, another for Interpolated Learning 2 output, and another for original learning recall. During each output the subject was allowed 2.5 minutes to answer as many questions as possible in the order of his own choice.

Immediately following original learning the experimental groups were given Interpolated Learning 1, then Interpolated Learning 2, and the control group was occupied with a nonrelated task of long division for the 9 minutes taken by the experimental groups for interpolated learning. Immediately following the interpolated activities each group was given the output questions from original learning for recall, and 2.5 minutes were given to answer as many questions as possible.

Results and Discussion

Original learning. The mean number of correct answers in original learning, as well as in interpolated learning and recall, is presented for each group in Table 1. The means of the four groups during original learning did not differ significantly; $F < 1.00$, $df = 3/76$. Since the groups learned the same passage during original learning, it may be concluded that they were comparable in learning ability and that a comparable degree of original learning was achieved. Consequently, these variables cannot confound recall in the four conditions.

Transfer. The major concern is with recall; however, it will be worthwhile to note briefly some transfer results. Although lacking the precise controls for transfer, some evidence may be obtained by examining learning over successive passages (original learning, Interpolated Learning 1, and Interpolated Learning 2) in the experimental groups. Table 1 shows that there was no divergence of the Different, Similar, and Identical conditions over successive passages; $F = 2.00$, $df = 4/114$ for the Conditions × Passages interaction. Thus, similarity did not differentially affect transfer over succesive passages. When Different, Similar, and Identical conditions were combined, learning increased slightly from original learning to Interpolated Learning 1 with little change from Interpolated Learning 1 to Interpolated Learning 2; $F = 3.59$, $df = 2/114$, $p < .05$ for the main effect of passages. While conditions of similarity between successive learning tasks typically produce negative transfer (cf. Underwood, 1966), the present findings do not rule out negative transfer effects in learning successive prose passages. They may mean only that unknown factors operating during learning are masking such effects. Such a situation is not uncommon in transfer (cf. Martin, 1965), especially if the negative effects are small.

TABLE 1

MEAN NUMBER OF CORRECT ANSWERS IN ORIGINAL LEARNING, INTERPOLATED LEARNING 1, INTERPOLATED LEARNING 2, AND RECALL

Condition	Original Learning	Interpolated Learning 1	Interpolated Learning 2	Recall
Control	12.75	—	—	12.30
Different	13.20	15.00	13.70	12.80
Similar	13.85	13.60	14.50	6.65
Identical	13.65	14.85	14.95	7.75

Recall. Table 1 shows that there was an overall retention decrement between original learning and recall; $F = 30.14$, $df = 1/76$, $p < .01$ for the main effect of retention interval, and that the retention decrement differed widely for the four conditions; $F = 7.92$, $df = 3/76$, $p < .01$ for the Conditions × Retention Interval interaction. A separate analysis of only the recall scores produced a significant F value between the four conditions; $F = 13.67$, $df = 3/76$, $p < .01$. Individual comparisons using the Newman-Keuls procedure indicated that there was no recall decrement in the Different condition relative to the control, but there was a significant decrement in both the Similar condition ($p <. 01$) and the Identical condition ($p < .01$). There was also a significant decrement in the Similar and Identical conditions relative to the Different condition (both $p < .01$). Recall in the Similar and the Identical conditions did not differ significantly.

These recall data clearly demonstrate that an increase in interpassage similarity between successive prose-learning tasks increases retroactive inhibition. They are in line, therefore, with the effects of increasing similarity of materials learned in successive tasks for free-recall, paired-associate, and serial learning (cf. Keppel, 1968). Relative to the control condition there was no recall decrement in the Different condition, a 45.93% decrement (5.65 answers) in the Similar condition, and a statistically equivalent 36.99% decrement (4.55 answers) in the Identical condition. Thus, the major effect of interpassage similarity was between the Different and the Similar conditions, with no additional effect between the Similar and the Identical conditions. This latter finding is not surprising in view of the three conditions of similarity, which might be thought of as involving greater similarity manipulation between the Different and the Similar conditions, than between the Similar and Identical conditions. That is, in the Different condition the successive passages were constructed so that the experimenter-generated questions were different, in the Similar condition the passages were constructed so that the questions were similar, and in the Identical condition they were constructed so that the questions were identical. The only change between the Similar and the Identical conditions was in the name of the hypothetical person.

EXPERIMENT II

It is possible that a substantial portion of the retroactive inhibition in Experiment I resulted from failures of passage differentiation; that is, the answers to corresponding original learning, Interpolated Learning 1 and Interpolated Learning 2 questions were not lost from memory, but rather the discrimination of which answer went with which question was lost. For example, consider the retroactive inhibition in the Similar condition. Corresponding questions from original learning and Interpolated Learning 1 were "Payton was born in what city?" and "Fowler was born in what city?" Both correct answers may be remembered, but not the memory for which answer goes with which question. Under conditions where recall of only original learning answers is taken, as in Experiment I, failures of passage differentiation may contribute to retroactive inhibition. In Experiment II, original learning recall was taken, and this was followed by Interpolated Learning 1, then Interpolated Learning 2 recall. If failures of passage differentiation contribute to retroactive inhibition, then a significant number of original learning answers should occur to corresponding Interpolated Learning 1 and Interpolated Learning 2 questions.

Method

Two conditions were employed, a control which duplicated the control condition in Experiment I, and an experimental condition. The experimental condition duplicated the Similar condition in Experiment I with the addition that following original learning recall the Interpolated Learning 1 questions were given with 2.5 minutes for recall, and then the Interpolated Learning 2 questions were given with 2.5 minutes for recall. The passages, questions, and question orders were the same ones used in Experiment I. The question order used for original learning recall was used also for Interpolated Learning 1 and Interpolated Learning 2 recall in the experimental condition. All other procedures were the same as in Experiment I. Thirty subjects, undergraduate education students at the University of Delaware, were assigned to each of the two conditions.

Results and Discussion

The mean number of correct answers given in original learning was 11.67 for the control condition and 12.33 for the Similar condition; $F < 1.00$, $df = 1/58$. Therefore, it may be assumed that the two groups were comparable in learning ability and that an equivalent degree of learning was achieved. Learning of the successive passages in the Similar condition is indicated by the number of correct answers in original learning, Interpolated Learning 1 and Interpolated Learning 2; these values were 12.33, 13.80, and 14.97, respectively, $F = 8.59$, $df = 2/58$, $p < .01$. As in Experiment I, there was some increase in performance over successive passages.

The mean number of correct answers for original learning recall in the control condition was 12.37 and the Similar condition, 7.73; $F = 25.69$, $df = 1/58$, $p < .01$). Consequently, the Similar condition again produced substantial retroactive inhibition. The question of major interest, however, is the number of original learning answers that were given during Interpolated Learning 1 and Interpolated Learning 2 recall. For each original learning question, the number of correct original learning answers which occurred to corresponding Interpolated Learning 1 and Interpolated Learning 2 questions during Interpolated Learning 1 and Interpolated Learning 2 recall was computed, and these values were summed for each subject. The mean value for the subjects in the Similar condition was .44. If these recalls are counted as correct in order to estimate correct original learning recall in the absence of any failures of differentiation, correct original learning recall becomes 8.17 (7.73 plus .44). This lenient measure of correct original learning recall is not significantly higher than the stringent measure of correct original learning recall in the Similar condition (7.73), $F = 1.16$ $df = 1/29$, suggesting that failures of differentiation do not contribute to retroactive inhibition. Finally, the amount of retroactive inhibition produced when the lenient measure of original learning recall is compared to the control is still highly significant; $F = 21.77$, $df = 1/58$, $p < .01$.

EXPERIMENT III

It is possible that the retroactive inhibition observed in both Experiments I and II was dependent upon the output questions which followed each passage. For example, it is well established that learning occurs during output (cf. Tulving, 1967), and it is possible that learning during output on successive passages was responsible for the retroactive inhibition observed. If this is the case, retroactive inhibition should not be found when output is eliminated following original learning, Interpolated Learning 1, and Interpolated Learning 2. Experiment III tested this possibility. In the experimental condition, original learning, Interpolated Learning 1, and Interpolated Learning 2 were learned in succession and then original learning recall was taken. In addition to original learning recall, Interpolated Learning 1 and Interpolated Learning 2 recall were also taken to determine the contribution of failure of passage differentiation to retroactive inhibition under these conditions.

Method

Two conditions were employed, a control for forgetting and an experimental condition that was similar to the Similar condition in Experiment II. The difference between the Similar condition in Experiment II and Experiment III was the elimination of output following original learning, Interpolated Learning 1, and Interpolated Learning 2. Therefore, in the Similar condition there was original learning input, then Interpolated Learning 1 input, then Interpolated Learning 2 input, each for 2 minutes. In the control condition there was original learning input for 2 minutes followed by a nonrelated task of long division for the 4 minutes required for interpolated learning in the Similar condition. Immediately following the interpolated activities, both groups were given original learning recall for 2.5 minutes. Following original learning recall in the Similar condition, Interpolated Learning 1 then Interpolated Learning 2 recall was given, each for 2.5 minutes. The passages used in both conditions were the same ones used in Experiment II, as were the questions used for original learning, Interpolated Learning 1, and Interpolated Learning 2 recall. The same ordering of questions used in Experiment II for original learning, Interpolated Learning 1, and Interpolated Learning 2 was also used in Experiment III. All other procedures were the same as in Experiment II. Thirty subjects, undergraduate education students at the University of Delaware, were assigned to each condition.

Results and Discussion

Since output did not follow original learning, equivalence of the two conditions in learning ability and degree of original learning must rest upon the assumptions of randomization. Learning of the successive passages in the Similar condition also cannot be assessed because of the absence of output following original learning, Interpolated Learning 1, and Interpolated Learning 2 inputs. The mean number of answers correct in original learning recall was 11.30 in the control condition and 8.37 in the Similar condition; $F = 11.93$, $df = 1/58$, $p < .01$. Therefore, retroactive inhibition was obtained when output following each passage was removed.

The amount of retroactive inhibition may be compared for the Similar condition in Experiment II and in Experiment III. The difference between the control and the Similar condition in Experiment II (12.37 versus 7.73) did not differ from the difference between the control and the Similar condition in Experiment III (11.30 versus 8.37); $F = 1.77$, $df = 1/116$, for the Control versus the Similar Condition × Experiment II versus Experiment III interaction. This suggests that the elimination of output in

Experiment III had no effect on the amount of retroactive inhibition; that is, retroactive inhibition is not dependent upon the process of output. These data, however, must be interpreted in the light of the shorter time interval between the end of original learning input and original learning recall in Experiment III (4 minutes as opposed to 9 minutes in Experiment II), as well as the fact that the comparisons were cross-experimental. However, neither of these facts would seem able to account for the equivalent amounts of retroactive inhibition.

In order to estimate correct original learning recall in the absence of any failures of passage differentiation, the number of correct original learning answers which occurred to corresponding Interpolated Learning 1 and Interpolated Learning 2 questions during Interpolated Learning 1 and Interpolated Learning 2 recall in the Similar condition was computed. The mean value was .40 which, when counted as correct recalls, produced a lenient original learning recall mean of 8.77 correct answers (8.37 plus .40). This lenient measure of original learning recall is not significantly higher than the stringent measure of original learning recall (8.37), $F = 2.11$, $df = 1/29$, suggesting that failures of passage differentiation did not contribute to retroactive inhibition. The amount of retroactive inhibition obtained when this lenient measure of original learning recall is compared with the control is still highly significant; $F = 8.64$, $df = 1/58$, $p < .01$.

Discussion

In summary, two major findings emerge from the present studies, and these may be highlighted. First, the results clearly indicate that interpassage similarity influences retroactive inhibition in prose-learning tasks. A number of recent studies have failed to find any retroaction effects in prose learning (Gaite, Ausubel, & Stager, 1969; Schuell & Hapkiewicz, 1969). It is possible that in these studies, an analysis of the successive passages in terms of experimenter-generated questions would reveal them to be very much like the passages in the Different condition of Experiment I in which no retroactive inhibition was found. Second, the results of the present investigations indicate that retroactive inhibition cannot be attributed simply to failures of passage differentiation or to output phenomena. If retroactive inhibition of prose-learning tasks is to be explained, additional mechanisms will be needed.

REFERENCES

Crouse, J. H. Transfer and retroaction in prose learning. *Journal of Educational Psychology,* 1970, **61,** 226–228.

Gaite, A. J. H., Ausubel, D. P., & Stager, M. Retroactive facilitation in meaningful verbal learning: Generality over time, and effect of differing levels of verbal ability. Paper presented at the meeting of the American Educational Research Association, Los Angeles, February 1969.

Keppel, G. Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.) *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J., Prentice-Hall, 1968.

Martin, E. Transfer of verbal paired associates. *Psychological Review,* 1965, **72,** 327–343.

Postman, L. Unlearning under conditions of successive interpolation. *Journal of Experimental Psychology,* 1965, **70,** 237–245.

Schuell, T. J., & Hapkiewicz, W. G. Similarity and retroaction in prose material. Paper presented at the meeting of the American Educational Research Association, Los Angeles, February 1969.

Tulving, E. The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6,** 175–184.

Underwood, B. J. *Experimental psychology.* (2nd ed.) New York: Appleton-Century-Crofts, 1966.

(Received November 27, 1969)

Journal of Educational Psychology
1971, Vol. 62, No. 1, 45–48

STIMULATION, ENJOYMENT, AND ORIGINALITY IN DYADIC CREATIVITY

E. PAUL TORRANCE[1]
University of Georgia

It was hypothesized that individuals taking a test of creative thinking in dyads would attain a higher level of originality and experience stronger feelings of stimulation, enjoyment, and originality of expression than individuals working alone under standard test conditions. The subjects were 100 college juniors and seniors (74 females and 26 males) and were assigned randomly to the experimental and control conditions. In the experimental condition subjects were instructed to sit together and to call out their responses as they wrote them, to hitchhike on one another's responses, but not to repeat one another's responses. The Ask-and-Guess Test of the Torrance Tests of Creative Thinking (Form A) was used as a warm-up task and the Product Improvement Test of this same battery was the test task. A set of 10-point rating scales was used to obtain self-ratings of feelings of stimulation, enjoyment, and originality of expression. The findings supported all four of the hypotheses.

The history of creative achievement as it has been written is largely the story of individual creativity. There have been of course the rare but well-known instances of dyadic creativity such as that of Wilbur and Orville Wright, Marie and Pierre Curie, Charles and William Mayo, William and Karl Menninger, and Frederick Banting and Charles Best. Less well known was Alexander Graham Bell's Thomas A. Watson; Robert H. Goddard's wife, Esther; Thomas Edison's John Krueski and other collaborators. In behavioral science there have been the well-known husband and wife teams such as Sheldon and Eleanor Glueck, and Lois and Gardner Murphy, Thelma and Edward L. Thurstone, Margaret and Harry Harlow, Anne Roe and George Gaylord Simpson, and others. Also familiar is the expression, "Behind every great man is a great woman." Perhaps the underlying dynamic of this truth springs from the dyadic creativity of the husband-wife interaction. Thus, dyadic creativity may be more common and more important than is apparent from the history of creative achievement.

Neither the literature of social psychology and group dynamics nor creativity has anything specific to say about dyadic creativity. One of the few psychologists who have attempted to describe the nature of dyadic creativity is Henry A. Murray (1964). He conceptualizes dyadic creativity as dealing with two interdependent regions of imagination operating as a single system. In the dyadic interaction, both members of the dyad try to translate their imaginations into actual, overt reciprocations and collaborations. Murray (1964) uses the analogy of "two people singing a duet and making up the music as they go along [p. 639]." He attempts to formulate the concept of dyadic creactivity in terms of different kinds of zestful and capable transmissions resulting in appreciative receptions. In this formulation, the member of the dyad who acts as the transmittor in one instance may function as the receptor in another.

The present investigation arose from the author's research in the broader area of group dynamics and creative functioning and his interest in classroom procedures that facilitate creative functioning. In a study with 5-year-old children he (Torrance, 1969) had found a greater willingness to undertake difficult tasks in dyads than in individual and large group conditions. In a pilot study with both 5-year-old children and college students he (Torrance, 1970) had found that a higher level of originality is attained on creative thinking tests by both members of a dyad than by individuals working alone. The investigation reported herein represents an attempt to explore further the dynamics of dyadic creativity with particular emphasis on the feelings of members of dyads about their own stimulation, enjoyment, and originality. It was hypothesized that individuals taking a test of creative thinking in dyads would attain a higher level of creativity and experience stronger feelings of stimulation, enjoyment, and originality of expression than individuals working alone under standard conditions.

---

[1] Requests for reprints should be sent to E. Paul Torrance, Department of Educational Psychology, The University of Georgia, Athens, Georgia 30601.

METHOD

Subjects

The subjects were 100 juniors and seniors enrolled in three classes of introductory educational psychology at the University of Georgia. They were about equally divided between majors in elementary education and secondary education; 74 were female and 26 were male. The subjects were assigned randomly to the experimental and control conditions.

Measures

The Ask-and-Guess Test of the Torrance Tests of Creative Thinking, Form A (Torrance, 1966) was used as the warm-up task and the Product Improvement Test of this same battery was used as the test task. This task calls for the subject to produce ideas for improving a stuffed toy elephant so that it will be more fun for children to play with and requires 10 minutes. Responses are scored according to a standardized scoring guide for fluency, flexibility, and originality. Extensive reliability and validity data are reported in the norms-technical manual.

A set of 10-point rating scales was used in obtaining self-ratings of feelings of stimulation, enjoyment, and originality of expression. To permit chi-square analyses the 10-point scales for stimulation and enjoyment were collapsed to 3-point scales (High, points 1 and 2; Moderate, 3 and 4; Low, 5 and above). The 10-point scale for self-ratings of originality was collapsed to a 4-point scale (Highly Original, points 1 and 2; Fairly Original, points 3 and 4; Not Very Original, points 5 and 6; Unoriginal and Commonplace, 7 and above).

Experimental Treatment

The subjects within each of the three classes were assigned to dyads and to standard, individual, or control conditions. The test booklets had been marked in advance to designate specific dyads and were shuffled thoroughly before being drawn by subjects. Members of dyads were instructed to sit side by side and to call out their responses to one another as they wrote them. They were encouraged to hitchhike on one another's responses and permit themselves to be sparked by one another. They were forbidden, however, to copy or repeat one another's responses.

To provide warm-up and practice in working in dyads, the Ask-and-Guess Test was administered. This test consists of three 5-minute tasks: asking questions about a picture, making guesses about the causes of the action depicted in the picture, and making guesses about the consequences of the action depicted. Instructions for the test task were as follows:

> In the middle of this page is a sketch of a stuffed toy elephant of the kind you can buy in most dime stores for about one to two dollars. It is about six inches tall and weighs about a half pound. In the spaces on this page and the next one, list the cleverest, most interesting and unusual ways you can think of for changing this toy elephant so that children will have more fun playing with it. Do not worry about how much the change would cost. Think only about what would make it more fun to play with as a toy.

The rating scales were administered immediately after the subjects completed the test task.

Data were obtained for 100 subjects on the test task but the rating scales of two of the experimental and two of the control subjects could not be used.

RESULTS

The means, standard deviations, and *F* ratios for the scores on fluency, flexibility, and originality on the Product Improvement Test under dyadic and standard, individual, or control conditions are reported in Table 1. These data indicate that the experimental treatment had a very strong effect

TABLE 1

MEANS, STANDARD DEVIATIONS, AND *F* RATIOS FOR SCORES ON FLUENCY, FLEXIBILITY, AND ORIGINALITY ON PRODUCT IMPROVEMENT TEST UNDER DYADIC AND INDIVIDUAL CONDITIONS

Measure	Dyads		Individual		*F*
	M	*SD*	*M*	*SD*	
Fluency	23.7	8.75	19.0	6.32	9.38*
Flexibility	10.0	2.63	7.6	2.35	22.27*
Originality	27.1	13.86	18.2	9.13	14.36*

Note.—$n = 50$ for each condition.
* $p < .01$.

TABLE 2

SELF-RATINGS OF FEELINGS OF STIMULATION AND ENJOYMENT ON CREATIVITY TEST TASK IN DYADS AND INDIVIDUAL CONDITION

Variable	Dyads			Individual			χ^2
	High	Moderate	Low	High	Moderate	Low	
Stimulation	34	12	2	25	17	6	4.20*
Enjoyment	39	7	2	27	13	8	7.56*

Note.—$n = 50$ for each condition.
* $p < .05$.

on the creative functioning of the subjects in this study as measured by the fluency, flexibility, and originality scores of the Product Improvement Test. The results for flexibility and originality are especially strong.

Table 2 presents the data to permit a comparison of the self-ratings of the experimental and control subjects on feelings of stimulation and enjoyment on the test task. It will be noted from these data that the subjects in the experimental or dyadic condition expressed greater feelings of stimulation and enjoyment than the subjects in the control or standard administration of the test.

The data on the self-ratings of originality on the test task are reported in Table 3 for the experimental and control subjects. From these data, it is clear that the subjects in dyads sensed in their performance on the test task a stronger feeling of having produced highly original responses than did subjects working alone as under standard testing conditions.

TABLE 3

SELF-RATINGS OF ORIGINALITY ON CREATIVITY TEST TASK IN DYADS AND INDIVIDUAL CONDITION

Rating	Dyads	Individual
Highly original	17	3
Fairly original	18	20
Not very original	10	15
Unoriginal, commonplace	3	10

Note.—$n = 50$ for each condition. $\chi^2 = 14.64$, $p < .001$.

DISCUSSION

The data yielded by the experiment described herein give strong support to the author's hypothesis that individuals taking a test of creative thinking in dyads would attain a higher level of functioning and experience stronger feelings of stimulation, enjoyment, and originality of expression than individuals working alone under standard test conditions. It should be recognized that these results were obtained from college students who have reasonably good skills in communicating with one another. If similar results could be obtained from disadvantaged children and young people and potential school dropouts, the implications for compensatory education for these groups would be quite promising. It is known from a variety of studies that students in these groups are deficient in verbal skills and perform rather poorly on verbal tests of the kind used in this study. Many observers (Reissman, 1962; Taba & Elkins, 1966), however, have noted that disadvantaged children and youth are quite articulate in their peer groups.

The rationale supported by the findings of this study offers promise as a guide in devising classroom procedures to facilitate a higher level of creative functioning in the classroom. The production of original responses places the thinker in a minority of one and being a minority of one makes a person uncomfortable psychologically and requires courage and considerable ego strength. This makes children and young people in formal classroom situations reluctant to express their original ideas even on paper. Since disadvantaged children and youth, potential school dropouts, and the like are especially lacking in self-esteem and weak in ego strength, they are especially vulnerable to classroom pressures against originality of expression. If dyadic interaction can provide the support and stimulation necessary to reduce the pressures against originality of expression, the use of dyadic arrangements in classrooms might possibly result in a higher level of mental functioning among students, especially among disadvantaged children and young people, potential school dropouts, and others who require such psychological support and stimulation. The data also suggest that greater attention might profitably be given the entire problem of dyadic creativity.

REFERENCES

MURRAY, H. A. Dyadic creations. In W. G. Bennis, E. H. Schein, D. E. Berlew, & F. I. Steele(Eds.) *Interpersonal dynamics.* Homewood, Ill.: Dorsey Press, 1964.

REISMAN, F. *The culturally deprived child.* New York: Harper & Row, 1962.

TABA, H., & ELKINS, D. *Teaching strategies for the culturally disadvantaged.* Chicago: Rand McNally, 1966.

TORRANCE, E. P. *Torrance Tests of Creative Thinking: Norms-technical manual.* Princeton, N. J.: Personnel Press, 1966.

TORRANCE, E. P. Peer influences on preschool children's willingness to try difficult tasks. *Journal of Psychology,* 1969, **72,** 189–194.

TORRANCE, E. P. Influence of dyadic interaction on creative functioning. *Psychological Reports,* 1970, **26,** 391–394.

(Received November 27, 1969)

Journal of Educational Psychology
1971, Vol. 62, No. 1, 49–54

EFFECTS OF TEACHER'S EXPECTANCY ON STUDENT-TEACHER INTERACTION[1]

MYRON ROTHBART[2], SUSAN DALFEN[3], AND ROBERT BARRETT
McGill University

Although recent research on the effects of teacher expectation suggests that teachers' expectancies can influence student performance, the mechanisms that mediate this relationship remain unclear. To investigate these mediating processes, a simulated classroom situation was constructed in which teacher trainees discussed academic material with four students, two of whom were arbitrarily designated as "lacking in intellectual potential." Measures were obtained of (*a*) the amount of attention given to the high-expectation and low-expectation students, (*b*) the amount of verbal and gestural encouragement (reinforcement) allocated to high-expectation and low-expectation students, (*c*) the amount of talking done by high- and low-expectation students, and (*d*) the teacher's evaluation of both groups of students. It was found that teachers were more attentive to high-expectation students, but directed the same amount of reinforcement to both groups. Students in the high-expectation group talked more, presumably as a consequence of the teacher's greater attention. Teachers evaluated the low-expectation students as less intelligent, as having less potential for future success, and as having a higher need for approval.

Recent research by Rosenthal and Jacobson (1968) suggests that teachers' expectations may have an important effect on a student's intellectual development. A randomly selected group of students who were described to their teachers as likely to show marked intellectual gains, evinced a significantly higher gain on a group-administered IQ test than did control subjects. Although this research has been fairly criticized (e.g., Jensen, 1969; Thorndike, 1968), it is nonetheless reasonable to retain the hypothesis that a teacher's expectations can influence at least some aspects of a student's performance (Beez, 1968).

One of the important unanswered questions in Rosenthal and Jacobson's research concerns the mechanism by which the teacher's expectations affected the student. One hypothesis, that the teacher spent more time with the "spurters," was rejected by the experimenters on the grounds that teachers rated themselves as spending the same amount of time with both "spurters" and "nonspurters." Since it is quite possible, however, that teachers would be reluctant to admit—to themselves or others—preferential treatment for better students, it is difficult to take these self-ratings at face value. The researchers were more prone to argue that the quality of the interaction, rather than the amount of interaction, was responsible for the observed changes in students' performance. Presumably, the teacher's implicit and/or explicit encouragement accounted for the student's improvement.

The present research attempts to clarify the process linking teachers' expectation with changes in the students' behavior. More specifically, How would the teacher in a classroom setting behave differently toward "bright" than toward "dull" students? Would the teacher spend more time with the "bright" students? Or would she reinforce and encourage them more than the "dull" students? Or both? By studying the teacher's behavior toward students under experimental conditions in a simulated classroom setting, an attempt was made to observe (*a*) the teacher's allocation of time between "bright" and "dull" students, (*b*) the amount of reinforcement (encouragement) directed toward the two groups, and

---

[1] This research was supported in part by a grant to the first author from the National Research Council of Canada (No. APA 0263). The authors wish to thank the Protestant School Board of Montreal for their helpful cooperation.

[2] Requests for reprints should be sent to Myron Rothbart, Department of Psychology, University of Oregon, Eugene, Oregon 97403.

[3] Now at Cornell University.

(*c*) the resulting verbal output of the "bright" and "dull" students.

METHOD

Female teacher trainees at McGill University were recruited for research on "student interaction in a classroom setting" and asked to supervise a discussion with four students drawn from a local high school. Each discussion group consisted of one teacher trainee and four high school students, two of the latter being arbitrarily designated as "lacking in intellectual potential." Videotape recordings of the discussion session were scored to determine (*a*) the amount of time the teacher attended to the "brights" versus the "dulls," (*b*) the number of positive and negative reinforcements distributed to the two subgroups, and (*c*) changes over time in the amount of talking by the "brights" and "dulls."

Subjects

Thirteen female seniors at McGill University, serving as teachers, and 27 male and 25 female students from Montreal High School participated in the experiment. The McGill University seniors ranged in age from 20–22, and the students ranged from 13–15 years.

Nine of the 13 McGill subjects were in the final year of the Bachelor of Education program and the remaining 4 were fourth-year arts majors interested in teaching as a career. All McGill subjects were volunteers who responded affirmatively to an appeal for people to serve as teachers as part of a research program on classroom student interaction. The high school subjects were volunteers who responded to an appeal for students to participate in research concerned with assessing teaching techniques. While the four subjects within each experimental session were randomly assigned to a high-expectancy or low-expectancy condition, an attempt was made to construct groups of four that were homogeneous with respect to IQ scores and teacher's personality ratings.

Apparatus and Observation Procedure

The teacher was seated at the head of a rectangular table with two students seated at either long side. The teacher faced a one-way mirror which concealed a video camera in an adjacent room. Videotapes were made of the teacher during each discussion session. As high-expectation students were grouped along one side of a table and low-expectation students along the other (with sides alternating from session to session), it was possible to determine whether the teacher was attending to the high- or low-expectation students by noting which side of the table she was facing.[4] Experimenters behind the one-way glass recorded by means of stopwatches and cumulative time recorders (*a*) the amount of time the teacher spent attending to the high- and low-expectation students, and (*b*) the amount of time the high- and low-expectation students spent talking.

Procedure

Each teacher arrived in the laboratory approximately ½ hour before the beginning of the experimental session, was shown the observation room where the experiment would take place, and was given a false verbal description of the nature and purpose of the experiment. This description stated that the experimenters were interested in "...the way in which [students] interact with one another in a classroom setting ... [and] particularly ... with the way in which they react to the material which they encounter in class, individually and in a group." The experimenter informed the subject that her role as a teacher would be to conduct a classroom-like discussion of a literary passage while experimenters studied the behavior of the students; in addition, the students would be informed that "...we are interested in studying different approaches to English literature." The tone and content of the experimenter's communications to the teacher were designed to convey the impression that she was in fact a confederate of the experimenters, and that the primary interest was in the behavior of the students rather than herself.

The specific details of the procedure were then presented to the teacher. She was to consider a literary passage with the students, first by asking specific questions of each student, and later through an open, free-for-all-type discussion.[5] During the first half of the 30-minute session, she was to begin by posing three consecutive questions to each of the four students, spending approximately 3 minutes with each student. She would be provided with a list of questions to use, and was instructed to follow the questions in order. This question period was to be followed by a general discussion period lasting 15 minutes in which the teacher and students were permitted a free, unstructured discussion.

The teacher was then allowed to read the passage and questions in private, while the students were seated around the table in the experimental room. The students were given name tags and assigned to specific seats. Each was given a copy of the literary passage along with written instructions which stated in essence:

1. The present research was concerned with assessing different approaches to teaching English literature.

2. The teacher would be directing questions to each student, and then would lead a discussion concerning the literary passage.

---

[4] About half way through the experiment, the video machine broke down, and it was necessary to make all observations directly through the one-way mirror, rather than off the videotape.

[5] The literary passage was a paraphrase of Jonathan Swift's "Meditation on a Broomstick."

3. The cooperation of each student in discussing the passage would be appreciated.

The teacher was brought into the room with the students, introduced to them, and seated at the head of the table. The experimenter then asked the students to study the passage for a few moments, and handed the teacher the following written instructions, ostensibly to refresh her memory:

> We have brought together four grade eight and nine students in order to observe the way in which they interact with one another in a classroom setting. We are particularly concerned with the way in which they react to the material which they encounter in class, individually and in a group. Your task will be to conduct a classroom-like discussion based on the literary passage which we have distributed. You will direct specific questions to each of the students in turn, 3 questions *in a row* to each one while the others listen or think about the passage. If after three minutes you have not asked three questions, finish up with that student and go on to the next one. The specific questions should therefore take about 12 minutes. Following this question and answer period, which should actually be a short dialogue between you and each student, you are to lead the students in a general discussion of the passage, using the questions which we have provided as a guide. As we wish to stimulate a real classroom discussion in this short time, we would ask you not to be afraid to praise, criticize, goad, or just react strongly to the students, as the case merits. The discussion should last about 15 minutes, but you can forget about the time as we will interrupt the session when 15 minutes are up.
>
> The students have been told that we are interested in studying different approaches to teaching English literature. They are to consider this an extension of English class and are to cooperate with you, the teacher, in discussing the passage. Class tests, IQ scores, and teacher-evaluation of these students indicate that ________ ________ and ________ have considerably greater academic ability than ________ and ________ ________ who lack intellectual potential.

The two high-expectation students were always seated at one side of the teacher with the low-expectation students on the other side, with sides being alternated from session to session.

Dependent Measures

Attention. A cumulative record was made of the total amount of time the teacher spent attending to (speaking to and/or looking at) the high-expectation and low-expectation students. Recordings were made by noting which side of the table she was attending to. During the first half of the experiment, when the video recorder was functioning properly, all time measurements were made off a video replay; during the second half of the experiment, all measurements were made directly in the observation room. Since the task of recording attention required little or no personal judgment, the interscorer agreement on this variable was virtually perfect.

Reinforcement. An attempt was made to measure the number of positive and negative reinforcements allocated to the high-expectation and low-expectation students by having two independent raters judge the teacher's behavior either off a video replay or through the one-way mirror. The two raters, unaware of the purpose or aim of the experiment, were asked to judge the teacher's behavior as if they were the objects of her remarks, and to record the number of positive and negative reinforcements, using the following guide:

Positive Reinforcement

Type	*Examples*
Short verbal expressions	Yes, uh-huh, right, correct, nice, exactly!, okay, good, etc.
Encouraging verbal statements	That is possible, that is an interesting argument, etc.
Facial expressions connoting attention and/or encouragement	Smile, nod, etc.
Verbal congruence, that is, expanding upon student's idea	You are suggesting that. . . ; this would imply that. . . ; etc.

Negative Reinforcement

Type	*Examples*
Short verbal expression	No, uh-uh, wrong, I don't think so, etc.
Discouraging verbal statement	That is unlikely, I don't follow that, etc.
Facial expressions connoting disinterest and/or disapproval	Frown, scowl, head nods, etc.
Impatience, that is, cutting off student without allowing completion of an idea	Let's go on to the next question, that's enough of this discussion, etc.

TABLE 1

TEACHERS' ATTENTION, REINFORCEMENT, AND EVALUATION OF HIGH- AND LOW-EXPECTATION STUDENTS

Students	Attention (Minutes)	Reinforcement	Attribution of traits							
			Intelligence	Cooperation	Appeal	Curiosity	Interest	Need for approval	Contribution	Future success
High expectation										
$\bar{X}$	6.68	18.39	5.71	6.44	6.13	5.20	5.56	4.15	5.51	6.05
SD	1.50	8.11	2.23	2.33	2.25	2.55	2.30	2.33	2.71	2.19
Low expectation										
$\bar{X}$	5.50	19.15	4.40	6.72	6.31	5.34	6.02	6.20	4.93	4.42
SD	1.62	8.95	2.06	2.13	1.14	1.71	1.55	1.28	2.67	1.75
t^a	2.09	.45	1.97	.46	.35	.20	.74	3.50	.77	2.79
p^b	$<.06$	—	$<.08$	—	—	—	—	$<.01$	—	$<.02$

[a] All *t* values were calculated for correlated samples.

[b] All significance tests are two-tailed.

It should be noted that the above "typology" was to be used only as a guide, and that the rater was not required to distinguish among the various subcategories of reinforcement indicated above. The rater's task was simply to record an instance of positive or negative reinforcement, and to indicate whether that reinforcement was directed toward the left or right side of the table. The two raters were trained on several practice videotapes until they were able to come to substantial independent agreement. During the actual experiment both raters independently scored each teacher's behavior in order to assess interrater agreement. A rank-order correlation between the first and second rater's ordering of the 13 teachers with respect to amount of emitted positive reinforcement (amount of negative reinforcement was negligible) yielded a value of $r = .83$ ($p < .01$) for the high-expectation condition and $r = .93$ ($p < .01$) for the low-expectation condition. Thus, interjudge agreement on the measure of reinforcement was quite good.

Student's verbal output. During the general discussion period (the second half of the experimental session), a quantitative measure of verbal output from the high- and low-expectation students was obtained.

Teacher's attribution of traits. Following the experimental session, the teacher was asked to rate each of the four students on the following attributes: intelligence, cooperation, appeal, curiosity, interest, need for approval, contribution to the discussion, and potential for future success. Each attribute was represented on a bipolar continuous rating scale.

RESULTS

Data obtained on teacher's attention, reinforcement, and evaluation of students are presented in Table 1.

Attention. During the 15 minute discussion period, the teachers spent more time attending to the high-expectation than the low-expectation students ($t = 2.09$, $df = 12$, $p < .06$, two-tailed test).[6]

Reinforcement. As indicated earlier, the amount of negative reinforcement directed toward either high- or low-expectation students was negligible; only the positive reinforcement data were analyzed. There was no significant difference in the average amount of positive reinforcement distributed to the high-expectation versus the low-expectation students ($t = .45$, $df = 12$, *ns*).

Student's verbal output. Data on student's verbal output was subjected to a three-way analysis of variance (Teachers × Expectation × Time). The tendency for high-expectation students to talk more than low-expectation students was marginally significant ($F = 4.34$, $df = 1/12$, $p < .06$), and there was no tendency for this difference to change over time (interaction $F = .29$, $df = 4/48$, *ns*).

Teacher's attribution of traits. Of the original eight traits rated by the teacher, three discriminated between the high- and low-expectation students. Teachers rated high-expectation students as somewhat more intelligent ($t = 1.97$, $p < .08$, two-tailed test) and as having greater potential for future success ($t = 2.79$, $p < .02$, two-tailed test); the low-expectation students were perceived as having higher need for approval ($t = 3.50$, $p < .01$, two-tailed test).

---

[6] Note that the unit of analysis is the teacher, rather than the student (hence the $df = 12$). One of the criticisms of the Rosenthal-Jacobson research was that the expectation effect had not been demonstrated *across* teachers, but only *within* teachers.

DISCUSSION

The present data suggest that teachers pay more attention to the "better" students, and that these students in turn respond by talking more. There was no evidence that teachers gave more verbal or gestural encouragements (positive reinforcement) to high-expectation than low-expectation students. While teachers rated the high-expectation students as more intelligent and as having greater potential for future success, it is likely that this judgment reflects the success of the experimental manipulation rather than any independent assessment of the student's level of competence.

The finding that low-expectation students were evaluated as having greater need for approval is open to at least two possible interpretations.[7] First, it may be that the stereotype of the student with poor academic potential is one that includes the trait of "high need for approval." This assumes that the teacher's evaluation on this dimension was independent of her interaction with the students and that the same judgments would have been obtained if she were simply asked to describe her impressions of the typical student with poor academic ability. The second interpretation assumes that the obtained rating was influenced by feedback obtained during the discussion sessions. If the teacher is clearly showing more interest in the high-expectation students, the low-expectation students may recognize this and attempt to bid for the teacher's approval. The teacher, in turn, correctly perceives the low-expectation students' behavior as reflecting high need for approval. If the second interpretation is correct, it describes a situation in which A bases his attribution of a trait to B on behavior that was originally evoked by A's expectations. While this process may not be an example of a self-fulfilling prophecy (since A does not consciously attribute the trait to B at the outset of the interaction), it nevertheless indicates the complex manner in which interpersonal feedback may influence the attribution process.

Two further comments are in order: One concerning a qualification of the present data, and another concerning a possible alternative interpretation of these results. The first comment concerns the measure of positive reinforcement. Although the data indicate that teachers did not allocate more positive reinforcement to the high-expectation students, it would nonetheless be premature to conclude that verbal and gestural encouragements do not serve as a medium for transmitting teacher expectation. The present technique for measuring emitted reinforcement, while enjoying high interrater reliability, has not been systematically validated, and at present can only be justified in terms of face validity. The only reasonable conclusion at present is that *this* measure of reinforcing behavior did not discriminate teachers' behavior toward high-expectation and low-expectation students.

A possible alternative interpretation of the present data is predicated on the assumption that teachers did selectively reinforce high-expectation students, but that our measure of positive reinforcement failed to pick this up. This interpretation states that during the predicussion period high-expectation students were encouraged more than low-expectation students and that this encouragement affected the students' differential output during the discussion. The finding that teachers more frequently attended to high-expectation students is accounted for by reversing the cause and effect sequence; that is, by assuming that teacher's attention was affected by students' talking rather than the other way around. While this interpretation cannot be rejected with certainty, two considerations reflect on the merits of this proposal. First, an analysis of emitted positive reinforcement during the prediscussion period *only* suggests again no difference in amount of reinforcement allocated to high- versus low-expectation students ($t = .18$, *ns*). Second, it was apparent from the videotapes and the live sessions that virtually all interchanges between teachers and students were initiated by the teachers rather than by the students. This suggests that teachers were not passively responding to students (as implied by the above interpretation), but were actively initiating the interaction.

---

[7] It should be noted that high-expectation students were rated approximately "neutral" on "need for approval," whereas low-expectation students were rated as having a high need for approval.

The results from the present experiment, then, support previous research indicating that teachers' expectations influence student performance. No evidence was obtained for teachers directing greater positive reinforcement toward the "better" students. Teachers did spend more time attending to the "better" students, and these students spent more time talking. As it stands, the present findings call into question Rosenthal and Jacobson's proposal that the effects of teacher expectation are not mediated by selective attention to the "better" students.

REFERENCES

BEEZ, W. V. Influence of biased psychological reports on teacher behavior and pupil performance. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 1968, **3**, 605-606.

JENSEN, A. R. Review of "Pygmalion in the classroom." *American Scientist*, 1969, **51**, 44A-45A.

ROSENTHAL, R., & JACOBSON, L. *Pygmalion in the classroom.* New York: Holt, Rinehart & Winston, 1968.

THORNDIKE, R. L. Review of "Pygmalion in the classroom." *American Educational Research Journal*, 1968, **5**, 708-711.

(Received December 3, 1969)

Journal of Educational Psychology
1971, Vol. 62, No. 1, 55–59

FAILURE TO INCREASE LEARNING USING THE TIME SAVED BY THE TIME COMPRESSION OF SPEECH[1]

THOMAS G. STICHT[2]

Human Resources Research Organization, Monterey, California

Time-compressed tape recordings were used to determine if presenting new information in the time saved by the compression process would improve overall performance on a criterion listening comprehension test. Seventy-five high- and 75 low-aptitude men (18–25 years) served as subjects. Results indicated that learning of new material presented in the time saved by the compression process did occur, but overall learning did not surpass that obtained by listening to the uncompressed, unelaborated material. Thus it remains to be demonstrated that more information can be learned in a given unit of time with moderate time compression of materials.

By means of an electromechanical process, the rate of speech of a tape-recorded message can be accelerated without producing the "chipmunk" sound, which results whenever a tape is played at a speed faster than the original recording speed. The process involves the periodic deletion of small temporal segments of the speech, while the remaining speech segments are connected to produce a continuous message. By means of this process, the time required to listen to a tape-recorded speech is decreased. Thus such speech is sometimes called time-compressed speech (cf. Foulke & Sticht, 1969, for an extended discussion of the time-compression process).

One of the intriguing aspects of the use of time-compressed speech is that more information can be presented in a given amount of time. For instance, if a message is time-compressed by 50%, it is possible to present the compressed version two times in the same amount of time required to present the uncompressed version once. A second alternative is that extra information may be presented in the time saved by the compression process.

In previous work (Sticht, 1969) time-compressed, tape-recorded messages were used to determine if listening to the compressed message twice, in the same amount of time required to listen to the uncompressed message once, would improve listening comprehension scores of high- and low-mental-aptitude men. The results indicated that, for both groups of subjects, listening twice to the compressed tapes improved comprehension over that obtained by listening to the compressed tapes only once. However, the comprehension of repeated time-compressed messages was not improved over that obtained with a single listening to the uncompressed messages. Similar results have been reported by Fairbanks, Guttman, and Miron (1957a); Jester and Travers (1967); Friedman, Graae, and Orr (1967); Woodcock and Clark (1968a, 1968b), and Hopkins (1969).

To date then, research seems to indicate that using the extra time resulting from the time compression of materials to simply repeat information is not likely to improve learning over what could be obtained by listening once to the uncompressed message presented within the "normal" range of speech rate. Furthermore, the work of Fairbanks, Guttman, and Miron (1957b) suggests that listening twice to the *uncompressed* message is not likely to produce very drastic improvements in comprehension—if any at all. Possibly the effectiveness of repeated time-compressed messages may be increased for subjects who are trained in listening to time-compressed speech, but there are no firm data to suggest this (cf. Friedman et al., 1967).

---

[1] The research reported in this paper was performed at HUMRRO Division No. 3, Monterey, California, under Department of the Army contract with The Human Resources Research Organization; the contents of this paper do not necessarily reflect official opinions of policies of the Department of the Army. Reproduction in whole or in part is permitted for any purpose of the Department of the Army.

[2] Requests for reprints should be sent to Thomas G. Sticht, Human Resources Research Organization, P. O. Box 5787, Presidio of Monterey, California 93940.

A second possibility which has been mentioned for improving the efficiency of learning by listening is to use the time saved by the compression of material to present additional information. Fairbanks et al. (1957a) used the time saved resulting from 30% (201 words per minute) compression of a message to emphasize certain portions of the message. As they pointed out, this amounts to trading temporal redundancy for verbal redundancy. Their results indicated that the reinforcing of certain parts of the selection did, indeed, increase the comprehension of the emphasized materials. However, this increase appeared to occur at the expense of the remaining, unemphasized content, for the comprehension of this material showed a highly significant decline. Thus, the overall comprehension score for the reinforced, compressed material was less than the overall score for the uncompressed material.

Fairbanks and his associates (1957a) suggested that emphasizing certain parts of the message may have led the subjects to assume that verbal redundancy meant "important to learn" and hence such emphasis may have selectively focused attention upon certain parts of the message, while diminishing attention to the remainder of the material. This suggests that if the time saved by the compression process was used to present additional, new information, perhaps an overall increase in the amount learned in a given unit of time might occur. The present research was performed to test this hypothesis.

METHOD

Subjects

One hundred and fifty Army inductees served as subjects. Seventy-five of the men had Armed Forces Qualification Test (AFQT) scores less than 30, forming a low-mental-aptitude group. The remaining 75 men had AFQT scores of 80 or above and formed a high-mental-aptitude group. In terms of intelligence, these groups represent men having IQs $\leqq$ 90 or below and 120 plus, respectively (Hedlund, 1959).

Materials and Procedures

The tape-recorded message used in this study was the "Roland" selection from the standardized listening passages prepared by Clark and Woodcock (1967). The message was prepared at a normal, uncompressed speech rate of 178 words per minute. The time required to listen to the entire, uncompressed message was 15 minutes. For the purposes of the present research, five modified versions of the recording were prepared to constitute five experimental conditions. In one version, the message was presented at the normal, uncompressed speech rate of 178 words per minute. However, the listening time was restricted to only 6 minutes 4 seconds. Thus, information relevant to only the first half of a 28-item criterion test was presented in this condition. By means of the time sampling compression method, two time-compressed versions of the foregoing message were prepared. One version was compressed by 36%, which produced a speech rate of 278 words per minute, and reduced the listening time from 6 minutes 4 seconds to 3 minutes 53 seconds. The second compressed tape was at 53% compression. This resulted in a speech rate of 378 words per minute, and reduced the listening time from 6 minutes 4 seconds to 2 minutes 52 seconds. Thus, three versions of a message were available having speech rates of 178, 278, and 378 words per minute and for which the time needed to listen to the message decreased from 6 minutes 4 seconds, in the case of normal speech, to 3 minutes 53 seconds at 278 words per minute and 2 minutes 52 seconds using speech of 378 words per minute. These tapes were used to assess the effects of increasing the speech rate upon the comprehension of a recorded message, while holding the message content constant.

Two additional tapes were prepared at 36% (278 words per minute) and 53% (378 words per minute) time compression. In these tapes, however, the total listening time was held constant at 6 minutes 4 seconds. Thus, these tapes presented the same information as the three tapes described above, *plus* additional information in the time saved by the compression process.

For the five experimental conditions represented by the five versions of the listening selections, comparable, independent groups of high- and low-aptitude subjects were assigned as listeners. Subjects were assigned in an unsystematic manner, as they became available, until all treatment cells were filled to an *N* of 15. The subjects listened to the tapes in an ordinary classroom. They were seated in a semicircle around a tape recorder adjusted to a "comfortable" listening level determined by the group. Subjects listened first to the recorded instructions on the standardized listening tapes. Then they listened to the "Roland" selection. Immediately after the presentation of the listening selection, Form A of the comprehension tests was administered both by reading and listening. This test contains 28 four-alternative multiple-choice questions. In the present research, the 6 minutes 4 seconds of listening time presented at a normal (178 words per minute) speech rate provided answers relevant to only the first 14 (50%) of the 28 test items. This was true also for the two compressed versions in which the speech rate was 278 or 378 words per minute and the listening times were 3 minutes 53 seconds and 2 minutes 52 seconds, respectively. For the compressed versions in which the listening time was held constant at 6 minutes 4 seconds, information relevant to both the first and second halves of the comprehension test was presented. In this case, information relevant to 23 (82%) of the test items was presented at 278 words per minute, and information relevant to all 28 (100%) test items was presented at 378 words per minute. Of primary interest was whether or not holding the listening time of the compressed selections equal to that of the uncompressed message and hence presenting more information relevant to the criterion test would result in an overall increase in scores on the criterion test.

RESULTS

The results of the study are summarized in Figure 1. In this figure, the unfilled symbols designate the conditions for which the listening time was constant at 6 minutes 4 seconds. The filled symbols indicate those conditions for which listening time was reduced. The square symbols are for the high-aptitude subjects and the round symbols are for the low-aptitude subjects. The abscissa indicates the rate of speech at which the message was presented, and the ordinate is the percent correct on the 28-item comprehension test.

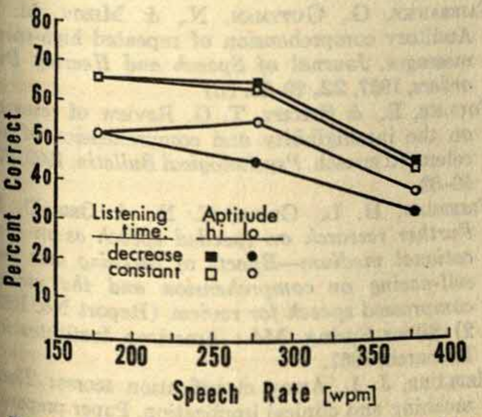


FIG. 1. Listening comprehension scores for high- and low-mental-aptitude subjects who listened to time compressed: time limited (filled symbols) or time compressed: time extended selections.

The data indicate that, under those conditions in which the speech rate was increased and the listening time was reduced, comprehension decreased for both high- and low-aptitude subjects. This is the typical finding regarding the relationship between speech compression and comprehension (Foulke & Sticht, 1969).

The data of primary interest are given by the unfilled symbols. In this case the listening time was constant at 6 minutes 4 seconds, while the speech rate was increased from 178 to 278 to 378 words per minute. Thus more information relevant to the criterion test was presented with the faster rates of speech. The data of Figure 1 indicate that, for both low- and high-aptitude men, there was no increase in comprehension scores over those for the uncompressed message when more information relevant to the test was presented at the accelerated speech rates. For lower aptitude men there is a suggestion that listening to additional information at accelerated speech rates may have improved comprehension over that obtained by listening to less information at the same accelerated speech rates. However, the differences indicated in Figure 1 are not statistically significant (*t* tests; $p > .05$).

DISCUSSION

These data indicate that listening to additional information in the time saved by the time compression of speech may not lead to increased learning. In the present study this was true even for material compressed only 36% and presented at 278 words per minute, and even though this compression resulted in a higher "listening efficiency" score, that is, more was learned per unit of time spent listening, than obtained with the normal (178 words per minute) rate of speech. This was so for both aptitude groups.

The present results are essentially the same as those found by Fairbanks et al. (1957a). But in the present case the possibility of selectively focusing attention through emphasis of materials was avoided; and hence does not explain the failure to find improved learning with extended listening. However, the notion of inhibition may be related to the present findings. An analysis of the responses of high-aptitude subjects to the first and second halves of the 28-item test indicated that, with the materials presented at 278 words per minute, the scores on the first half of the test decreased slightly when the message was 6 minutes 4 seconds in duration as opposed to when the message length was only 3 minutes 53 seconds in duration. Thus there is the possibility that retroactive inhibition may have occurred such that listening to additional material may have interfered with the retention of previously presented material. However, the evidence for this is very slight. Also this interpretation is not confirmed by the data for the low-aptitude men who, in fact, showed a slight increase in performance for both halves of the test when listening time at 278 words per minute was extended from 3 minutes 53 seconds to 6 minutes 4 seconds.

On the basis of these data and the work of others (Fairbanks et al., 1957a, 1957b; Friedman et al., 1967; Woodcock & Clark, 1968a, 1968b; Hopkins, 1969), it appears as though the technique of trading time for information has not resulted in more information being processed by the listener for short-term retention. Most significantly, this has been true for materials compressed to speech rates of 275–300 words per minute for which listening "efficiency," that is, the amount learned per unit of listening time, has actually been higher than obtained with "normal" materials. Thus, the implication that, because of improved listening efficiency more information can be learned in a unit of time with moderate compression, has yet to be substantiated.

It should be mentioned, however, that there are several features of the various research efforts under consideration which serve to limit conclusions to be drawn from them. For one thing, practically all of this research has involved listeners untrained in listening to compressed speech. Secondly, all of the studies have used materials within a given subject matter area. Possibly, the probability of interference factors might be reduced if a different type of content was presented in the time saved by the compression process. Thirdly, these studies have presented the additional information in a single sitting and immediately tested for learning. Perhaps some spacing of the presentation of new compressed information might increase learning over that obtained by continuously listening for the same amount of time to uncompressed materials (but see Friedman et al., 1967, and Woodcock & Clark, 1968a, 1968b, for negative findings using long-term intervals between repetitions of materials).

Finally, it should be pointed out that the studies reported in this paper have all been concerned with using the time saved by the time-compression process for increasing the learning of a given group of subjects. An alternative would be to use this time-savings for other purposes, such as instructing additional students. Many studies, including the present one, have demonstrated that much learning can occur with materials that have been compressed by 30–40%. Clearly this time savings can be used to instruct additional listeners. Thus, the efficiency of time-compressed listening does not rest solely on demonstrating an increase in the amount of learning per group per unit of time, but also by the demonstration that more groups per unit of time can be instructed with moderate compression.

REFERENCES

CLARK, C., & WOODCOCK, R. W. *Standardized listening passages.* (IMRID Report No. 2, Vol. IV) Nashville, Tenn.: George Peabody College, 1967.

FAIRBANKS, G., GUTTMAN, N., & MIRON, M. S. Auditory comprehension in relation to listening rate and selective verbal redundancy. *Journal of Speech and Hearing Disorders,* 1957, **22,** 23–32. (a)

FAIRBANKS, G., GUTTMAN, N., & MIRON, M. S. Auditory comprehension of repeated high-speed messages. *Journal of Speech and Hearing Disorders,* 1957, **22,** 20–22. (b)

FOULKE, E., & STICHT, T. G. Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin,* 1969, **72,** 50–62.

FRIEDMAN, H. L., GRAAE, C. N., & ORR, D. B. *Further research on speeded speech as an educational medium—Effects of listening aids and self-pacing on comprehension and the use of compressed speech for review.* (Report No. R67-2) Silver Spring, Md.: American Institutes for Research, 1967.

HEDLUND, J. L. Army classification scores: Their meaning and clinical implication. Paper prepared for the Conference of Army Clinical Psychologists, Washington, D. C., September 1969.

HOPKINS, F. W. An experimental study of the effects of repetitive compressed speech on listening comprehension. Unpublished doctoral dissertation, University of Maryland, 1969.

JESTER, R. E., & TRAVERS, R. M. W. The effect of various presentation patterns on the comprehension of speeded speech. *American Educational Research Journal*, 1967, **4,** 353–360.

STICHT, T. G. Comprehension of repeated time-compressed recordings. *Journal of Experimental Education*, 1969, **37,** 60–62.

WOODCOCK, R. W., & CLARK, C. *Comprehension of a narrative passage by fifth grade children as a function of listening rate and presentation strategy.* (IMRID Report No. 3, Vol. V) Nashville, Tenn.: George Peabody College, 1968. (a)

WOODCOCK, R. W., & CLARK, C. *Differences in learning through compressed speech as a function of presentation strategy and rate among culturally disadvantaged fourth grade children.* (IMRID Report No. 9, Vol. V) Nashville, Tenn.: George Peabody College, 1968. (b).

(Received December 11, 1969)

Journal of Educational Psychology
1971, Vol 62, No. 1, 60-66

WEBER IN THE CLASSROOM

LEONARD V. GORDON[1]
State University of New York at Albany

An instrument designed to measure the "bureaucratic personality" at the adult level, and based on the Weberian model, had been previously developed. The present paper describes an attempt to measure the same personality construct from the sixth- through the twelfth-grade level, with the school serving as the environmental frame of reference. The adult (WEPS) and student (SEPS) forms were found to be constructually comparable. The SEPS was significantly related to other personality variables in directions congruent with bureaucratic theory. Bureaucratic attitudes as measured by the SEPS decreased monotonically with increasing grade level, and were related negatively to intelligence and academic achievement and positively to favorableness toward a traditional school environment. Similar findings were obtained in India and Japan.

A great deal of concern has been expressed regarding the bureaucratization of the individual in present day society. Weber (1946), the father of bureaucratic theory, observed that the individual was becoming a cog in the machinery of modern bureaucracy and Blau and Scott (1962) considered this development to be the greatest threat to our liberty. Huxley (1946), Orwell (1949), and others have projected fictionalized accounts of the final submission of the individual to the controlling powers, while Fromm (1941) proposed that such an outcome might not be unwelcome to most of the victims. Interest in the nature and development of the bureaucratic personality, the impact of bureaucracy on the personality of the individual, and characteristics of individuals who are attracted to—or easily adapt to—this type of environment has been widespread but related research has been limited.

Research on problems such as those posed above is facilitated by the generation of relevant constructs and the development of instrumentation for their measurement. To this end, the author (Gordon, 1968a) postulated a personality construct "bureaucratic orientation," which is defined as a commitment to the set of attitudes, values, and behaviors that are characteristically fostered and rewarded in highly bureaucratized work environments. An instrument, the Work Environment Preference Schedule or WEPS (Gordon, 1968c) then was developed to measure this construct in adults and older students.

The WEPS yielded highly meaningful results in studies involving individuals who have openly resisted bureaucratic pressures or who have rejected highly bureaucratic work environments. It discriminated significantly between militant teachers and those who refused to strike (Khalique, 1969), army career and noncareer motivated United States Military Academy graduates (Bridges, 1967), and Air Force ROTC students who dropped out and those who remained (Bronzo, 1966). The WEPS also was found to be related to a variety of personality characteristics in directions congruent both with bureaucratic theory and prior research findings.

If, in fact, bureaucratic orientation represents a meaningful construct for the older adolescent it would be expected to be meaningful for the younger student as well, since personality development is a continuous process. The measurement of bureaucratic orientation in the earlier years would facilitate the study of its development, its correlates, and its relationship to the socialization process more broadly conceived. Unfortunately, the WEPS is not suitable for this purpose, since the language level of the items is too difficult for the typical grade school student and the concept of "organization" to which many of the items refer is unfamiliar.

---

[1] Requests for reprints should be sent to Leonard V. Gordon, Program for Behavioral Research, School of Education, State University of New York, 1400 Washington Avenue, Albany, New York.

What follows is an attempt to bring Weber into the classroom—that is, to develop a measure of bureaucratic orientation appropriate for the younger student, with the familiar school setting serving as the environmental frame of reference. The development of a school form and the presentation of data bearing on its validity are the subjects of this paper.

DEVELOPMENT OF THE ORIGINAL FORM

Since the school form was directly derived from the adult form, a brief description of the development of the latter will be presented as necessary background. The first step in the measurement of bureaucratic orientation required that the particular classes of attitudes, values, behaviors, etc., that are important for adapting to bureaucratic environments be specified. To this end the original descriptive schema of Max Weber (1946), which identifies the common denominator characteristics of most modern bureaucratic organizations, was adopted as a point of departure. These characteristics include a pyramidal network of supervisor-subordinate relationships, a high degree of specialization, a reliance on expert judgment, an impersonal social structure, a comprehensive set of rules and standard operating procedures, and a career-oriented membership.

A parallel schema, but describing *individual* rather than organizational characteristics, was developed. These characteristics consisted of a willingness to subordinate oneself to a superior, to restrict oneself to a narrowly prescribed area of activity, to have full confidence in the judgment of experts, to maintain impersonal on-the-job relationships, to follow rules and regulations exactly, and to readily accept the organizational and in-group norm.

A set of items was prepared for the six categories comprising the schema. A cyclical program of test administration, item analysis, and test revision was then undertaken, using a varied array of student and adult samples. Items reflecting the individual's willingness to restrict himself to a narrow range of activities and to have full confidence in expert judgment were found to be unrelated or negatively related to items of the other categories. Accordingly, these two categories were dropped from further consideration. From each of the four remaining categories the six items which had the best statistics were selected to comprise the final form of the WEPS.

DEVELOPMENT OF THE SCHOOL FORM

The school form was intended to measure the same construct as that measured by the WEPS. Thus, five of the original categories were modified, as required, so as to be relevant to the school setting. The sixth, which referred to the restriction of one's activities, was inappropriate to the school setting. Following are the resultant category descriptions:

(S) *Self-Subordination* . . . a desire to comply with the wishes of and to please one's teacher (in lieu of one's superior).

(U) *Uncriticalness* . . . an uncritical acceptance of the opinions of teachers (in lieu of experts or subject matter specialists).

(I) *Impersonalization* . . . a preference for impersonal or formal relationships in school (in lieu of on the job), particularly with one's teachers (in lieu of superiors).

(R) *Rule Conformity* . . . a desire for the security that following rules and regulations affords.

(T) *Traditionalism* . . . a need to identify with one's school (in lieu of one's organization) and to conform to the peer-group norm.

With these categories as a guide, 60 items, parallel to those in the first developmental form of the WEPS, but referring to the school rather than the work environment, were prepared. All items were worded so that agreement would reflect acceptance of the traditional school norm. No attempt was made to control for acquiescence by item reversal since acquiescence was considered to be a characteristic of bureaucratic orientation. The work of Dale and Chall (1948), the judgment of reading specialists, and a trial administration served to assure that the items were comprehensible at least at the sixth-grade level. As in the WEPS, the following five response categories were used: Strongly Agree, Agree, Undecided, Disagree, and Strongly Disagree.

The items then were administered to several samples of students in the sixth through the twelfth grades for item analysis purposes. Items relating to the Impersonalization category were found to have little variance in the school setting and were disagreed with by a very large majority of the respondents. Apparently students, unlike their adult counterparts, were disinclined to have formal or impersonal relationships in their "work" environments. Accordingly, this category was excluded from further consideration. On the other hand, in contrast to the findings obtained during the development of the adult form, "Uncriticalness" items were positively related to those in the remaining categories for all grade levels and were largely acceptable on a statistical basis. The six items that had the most satisfactory internal statistics were selected from Uncriticalness and from each of the three remaining categories to comprise the final instrument. This 24-item form was given the title School Environment Preference Schedule or SEPS (Gordon, 1968b).[2] Two specimen items from each category are as follows:

1. A student should always do what his teacher wants him to. (S)
2. A teacher should know the answer to any question in the subject he teaches. (U)
3. In school, a student should never do anything that might be considered wrong. (R)
4. What a student says about his school should agree with what most of the other students say. (T)
5. A student's first real loyalty in school is to his teachers. (S)
6. You can always get the right answer to any question if you know who to ask. (U)
7. One of the most important reasons for being in school is to learn how to obey and follow rules. (R)
8. When asked, a student should always say that his school is the best. (T)

The SEPS, like all Likert scales, is subject to the extremeness response set, that is, the tendency of the individual to mark the extreme response alternative whether agreeing or disagreeing. Differential weights which nullified this response set at the agreement but not the disagreement end of the continuum had been found to enhance the validity of the WEPS (Gordon, 1967b, see also Gordon, 1971). This same scoring scheme (two for Strongly Agree and Agree, one for Undecided and Disagree, and zero for Strongly Disagree) was found to be effective with the SEPS as well (see Gordon & Kikuchi, 1970), and accordingly was adopted.

The test-retest reliability of the SEPS based on a sample of 77 eighth-grade students and a 9-week interval between administrations was .75; a test-retest reliability of .62 was found for a sample of 136 eleventh-grade students when retested 1 year later as seniors.

COMPARABILITY OF THE TWO FORMS

If in fact the SEPS and WEPS do measure a comparable construct at the upper grade levels, they would be expected to be highly related to one another and to bear similar relationships to construct-relevant external measures. The following data are pertinent in this regard.

The SEPS and WEPS were administered in that order to a sample consisting of 70 male and 55 female senior high school students. Separate category scores as well as total scores were computed for both instruments to permit a more detailed comparison. It will be noted (Table 1) that the intercorrelations between category scores are highly uniform. When the respective category reliabilities are taken into consideration a single factor would appear to account for almost all of the reliable variance in both the WEPS and SEPS. The correlation between the total scores on the two instruments is found to be .95 when corrected for attenuation due to unreliability. These results suggest that the two instruments are measuring a highly comparable construct and parenthetically support the use of a single total score for each. Correlations between the SEPS and WEPS based on data from tenth- and eleventh-grade samples had a median uncorrected value of .65. This lower relationship may be due to the lesser relevance and consequently lower reliability of the WEPS at these grade levels.

---

[2] A copy of the SEPS, which includes directions to the student, will be sent upon request.

It had been previously noted in the development of the SEPS that the Impersonalization items which referred to the school setting had little variance and were largely disagreed with by students at all grade levels. On the other hand, it was found in the present analysis that the Impersonalization items in the WEPS were frequently agreed with by the students and also were substantially related to those in the four SEPS categories (Table 1). Thus, Impersonalization appears to have a different meaning to the student when the school and the organization serve as the environmental referrent. That the expressed *level* of bureaucratic orientation may well be a function of the environmental frame of reference employed in its measurement is more generally supported by the finding at all three grade levels, that, on the average, the students marked about 60% more agreement responses on the WEPS than on the SEPS, even though they tended to achieve a similar relative status on both.

TABLE 1

INTERCORRELATIONS AMONG THE "CATEGORY" AND TOTAL SCORES OF THE SCHOOL ENVIRONMENT PREFERENCE SCHEDULE (SEPS) AND THE WORK ENVIRONMENT PREFERENCE SCHEDULE (WEPS)

WEPS	SEPS				
	S	U	R	T	Total
Self-Subordination (S)	.68	.56	.66	.62	.73
Impersonalization (I)	.50	.52	.52	.56	.60
Rule Conformity (R)	.60	.58	.67	.59	.70
Traditionalism (T)	.55	.55	.60	.66	.68
Total	.66	.66	.68	.69	.78

Note.—Common categories are underlined. "U" represents Uncriticalness.

Merton (1940), Weber (1946), and others have proposed that work environments represented by the bureaucratic organizational model would appeal to individuals with certain specified personality characteristics. Accordingly, as part of the construct validation of the WEPS (Gordon, 1968a, 1970) relationships with a variety of personality measures were obtained. For purposes of assessing the validity of the SEPS, as well as its constructual similarity to the WEPS, correlations of the SEPS with most of these same personality variables were determined, based on tenth-, eleventh-, and twelfth-grade student samples.

Correlations between the WEPS and the values measured by the Survey of Interpersonal Values (Gordon, 1960) and the Survey of Personal Values (Gordon, 1967a) had been obtained for five adult samples and highly consistent, significant correlations had been found for all samples with four of the value scales—in a positive direction with Conformity and Orderliness and in a negative direction with Independence and Variety. Similar relationships may be noted between the SEPS and these same scales for the high school samples (Table 2). Like adults, students who score high on bureaucratic orientation are inclined to place a high value on conformist behavior and on being systematic and orderly, and a low value on having personal independence of action and on engaging in new or varied activities. These results, consistently obtained at both the adult and student levels, are in full accord both with prior findings regarding need satisfaction in bureaucratic environments and with inferences derivable from bureaucratic theory.

Additionally, significant relationships had been previously noted between the WEPS and six other value dimensions for at least two of the five adult samples tested, positively with Recognition, Practical Mindedness, and Goal Orientation and negatively with Benevolence, Leadership, and Achievement. Significant relationships in a corresponding direction occur between the

TABLE 2

CORRELATIONS BETWEEN THE SCHOOL ENVIRONMENT PREFERENCE SCHEDULE AND VALUES MEASURED BY THE SURVEY OF INTERPERSONAL VALUES AND SURVEY OF PERSONAL VALUES

Values	Grade		
	12th^a	11th^b	10th^c
Interpersonal			
Support	.10	−.08	.09
Conformity	.22*	.27**	.35**
Recognition	.03	−.01	.33**
Independence	−.34**	−.13	−.24*
Benevolence	−.08	−.18	−.21*
Leadership	.01	−.13	−.18
Personal			
Practical Mindedness	.14	.09	.28*
Achievement	−.10	−.03	−.14
Variety	−.22*	−.34**	−.21*
Decisiveness	−.29**	−.06	−.41**
Orderliness	.21*	.20*	.34**
Goal Orientation	.03	.23**	.18

a $N =$ [illegible].
b $N =$ [illegible].
c $N =$ [illegible].
* $p < .05$, one-tailed test.
** $p < .01$, one-tailed test.

SEPS and four of these scales for one of the three high school groups (Table 2). Also, Decisiveness which was negatively related to the WEPS for all five adult samples is significantly related to the SEPS in a negative direction for two of the three school groups. Thus, the tendency of the bureaucratic adult to place a higher value on being treated as important, on taking care of his possessions, and on having clearly defined work objectives, and to place a lower value on treating other people with consideration and on coming to quick decisions is found in embryonic form in his younger counterpart. On the whole, these characteristics fit the opprobrious stereotype of the older bureaucrat, and reflect some of the dysfunctional aspects of bureaucratic behavior described by Merton (1940).

Theoretically, it is expected that bureaucratic orientation would be positively associated with authoritarianism, dogmatism, and acquiescence, characteristics which are prevalent in bureaucratic environments. Fairly substantial, significant relationships in the predicted direction will be noted for the SEPS with Authoritarianism, as measured by the California F scale (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950) and with Rokeach's (1960) Dogmatism scale (Table 3), for all three school samples. Correlations with the Couch and Keniston (1960) measure of acquiescence, while in the expected direction, are considerably lower and significant in only one instance. The magnitudes of the above relationships are highly similar to those previously obtained for a college sample between the WEPS and Authoritarianism, Dogmatism, and Acquiescence, the respective values being .59, .45, and .20 (Gordon, 1970).

That the expression of bureaucratic attitudes does not appear to be predominantly socially undesirable or desirable, at least for the particular population sampled, is reflected in the generally low correlations between the SEPS and the Crowne-Marlowe (1964) Social Desirability scale (Table 3). The average correlation of .13 is comparable to the .07 previously found between this scale and the WEPS for adults.

TABLE 3

CORRELATIONS BETWEEN THE SCHOOL ENVIRONMENT PREFERENCE SCHEDULE AND OTHER PERSONALITY MEASURES

Measure	Grade		
	12th^a	11th^b	10th^c
Authoritarianism (California F scale)	.45**	.48**	.52**
Dogmatism (Rokeach)	.38**	.39**	.32**
Acquiescence (Couch & Keniston)	.10	.22**	.18
Social Desirability (Crowne & Marlowe)	.07	.08	.16*

a $N =$ [illegible].
b $N =$ [illegible].
c $N =$ [illegible].
* $p < .05$, one-tailed test.
** $p < .01$, one-tailed test.

FURTHER EVALUATION OF THE SCHOOL FORM

The SEPS and WEPS appear to be constructually comparable as evidenced by their high degree of interrelationships and by the similarity of their correlations with relevant external measures. The validity of the SEPS will next be examined in terms of its relationship to several school-related variables.

As the student matures he is given a larger measure of freedom and is expected to assume increasing responsibility for his own behavior. Concurrent with this reduction in environmental constraints is an increasing desire for independence on his part. These changes in organizational expectations and individual needs would be expected to be mirrored in a decrease in SEPS means with increasing grade level. That this is the case will be noted in Table 4 which presents means by sex and grade level for the entire student body at a small school which had a negligible dropout rate. A monotonic decrease in SEPS means occurs from the seventh through the twelfth grades both for the boys and girls.

It would be anticipated that students who are higher on bureaucratic orientation will have the more favorable attitudes toward an authoritarian or traditionalistic school environment. This hypothesis was tested by having the tenth-, eleventh-, and twelfth-grade students who participated in the earlier validation study evaluate their school 1 week later using the following 7-point scales: Boring-Interesting, Important-Unimportant, and Meaningless-Meaningful. The scales were scored so that high scores would represent a favorable attitude toward their traditionalistic school. Correlations between the SEPS and the summed attitude ratings for the three grade levels were found to be .44, .44, and .27, respectively, all significant beyond the 5% level.

Correlations of the SEPS with measures of "intelligence" and/or school grades based on data from schools in four different systems in upstate New York are presented in Table 5. Relationships with both types of intellective indexes are uniformly negative, the more bureaucratically oriented the student, the less bright is he likely to be and the more poorly is he likely to do in school. These results may indicate that the less bright student compensates by a greater adherence to what he considers to be institutionally acceptable behavior or that the student who relies on his superiors or peer-group members to make decisions or do his thinking for him is less able to cope with situations in which he must function intellectively on an individual basis.

TABLE 4

SCHOOL ENVIRONMENT PREFERENCE SCHEDULE MEANS AND STANDARD DEVIATIONS FOR STUDENTS IN SIX GRADES

Grade	Male			Female		
	n	*M*	*SD*	*n*	*M*	*SD*
7	30	31.3	6.0	35	27.7	5.2
8	33	26.7	6.3	35	26.6	5.5
9	34	25.9	5.8	35	24.7	6.5
10	36	25.2	7.1	36	24.7	6.5
11	30	22.1	5.6	32	22.7	6.6
12	34	21.9	7.2	36	19.2	6.5

TABLE 5

CORRELATIONS OF THE SCHOOL ENVIRONMENT PREFERENCE SCHEDULE WITH SCHOOL GRADES AND MEASURES OF SCHOLASTIC APTITUDE

School	Grade	Sex	*n*	Grades	Aptitude
A	12	Male	47	−.41**	
		Female	49	−.31	
B	8	Male	113	−.21*	−.[illegible]
		Female	114	−.28**	−.[illegible]
C	11-12	Male	[illegible]		−.[illegible]
		Female	[illegible]		−.[illegible]
D	8	Male	[illegible]		−.[illegible]
		Female	40		−.11

[a] California test of mental maturity.
[b] Preliminary SAT-Verbal.
$* p < .05$.
$** p < .01$.

DISCUSSION

Present evidence indicates that "bureaucratic orientation," as originally defined at the adult level, may be satisfactorily measured in the younger student, with the school setting serving as the environmental frame of reference. The adult and student measures are found to be highly intercorrelated and to bear similar relationships to a variety of external measures relevant to their construct validity. Additionally, as would be expected, the school form shows a mean decrease with increasing maturity, a positive correlation with favorableness toward an authoritarian school climate, and negative relationships with both intelligence and school performance.[8]

---

[8] Similar results have been noted in research in other cultures. For example, in India (Kakkar, 1969), monotonic decreases in SEPS means from Grades 7 through 11 have been reported for both male and female samples. In Japan (Kikuchi & Gordon, 1970) a decrease in SEPS means with increasing grade level was also found. In Grades 7 through 9, in separate analyses for boys and girls, negative correlations with IQ and school achieve-

The SEPS should prove to be a useful instrument for a number of research purposes. For example, the SEPS should be relevant in studies of the adaptability of the student to different types of school climates and of the effect of such environments on the personality of the student. The student's level of bureaucratic orientation would be expected to be differentially related to his performance in, and adjustment to, progressive, traditional, parochial, and military schools. Exposure to particular school environments or "change-oriented" curricula, such as those designed to foster self-reliance, openness, or critical thinking, or those in which discipline and obedience are emphasized, might well result in changes in bureaucratic orientation. It would be reasonable to hypothesize further that personality dimensions which are most highly correlated with the SEPS would be the more susceptible to modification in such programs.

Finally, bureaucratic orientation would appear to be an important personality construct in its own right. Its ontogenesis, its behavioral correlates, and its relationship to cognitive and other personality variables at various age levels would represent worthwhile subjects of study.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

BLAU, P. M., & SCOTT, W. R. *Formal organizations*. San Francisco: Chandler, 1962.

BRIDGES, C. *Prediction of attrition at the U. S. Military Academy*. West Point: U. S. Military Academy, Research Division, 1967.

BRONZO, A. F., JR. Discriminant analysis of attrition of AFROTC cadets. Unpublished master's thesis, Boston College, 1966.

COUCH, A. S., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, 1960, **60**, 151–174.

CROWNE, D. P., & MARLOWE, D. *The approval motive*. New York: Wiley, 1964.

DALE, E., & CHALL, J. S. A formula for predicting readability. *Educational Research Bulletin*, 1948, **27**, 11–20.

FROMM, E. *Escape from freedom*. New York: Farrar & Rinehart, 1941.

GORDON, L. V. *Survey of Interpersonal Values*. Chicago: Science Research Associates, 1960.

GORDON, L. V. *Survey of Personal Values*. Chicago: Science Research Associates, 1967. (a)

GORDON, L. V. Validity of scoring methods for bipolar scales. *Educational and Psychological Measurement*, 1967, **27**, 1099–1106. (b)

GORDON, L. V. Correlates of bureaucratic orientation. *Proceedings of XVIth International Congress of Applied Psychology*. Amsterdam: Swets & Zeitlinger, 1968. (a)

GORDON, L. V. *School Environment Preference Schedule—SEPS*. (Program for Behavioral Research)Albany: State University of New York, 1968. (b)

GORDON, L. V. *Work Environment Preference Schedule—WEPS*. (Program for Behavioral Research) Albany: State University of New York, 1968. (c)

GORDON, L. V. The measurement of bureaucratic orientation. *Personnel Psychology*, 1970, **23**, 1–11.

GORDON, L. V. Are there two extremeness response sets? *Educational and Psychological Measurement*, 1971, in press.

GORDON, L. V., & KIKUCHI, A. Response sets of Japanese and American students. *Journal of Social Psychology*, 1970, **82**, 143–148.

HUXLEY, A. L. *Brave new world*. New York: Harper, 1946.

KAKKAR, S. B. Bureaucratic orientation of Indian students. Unpublished paper, State College of Education, Patiala, 1969.

KHALIQUE, N. A study of bureaucratic orientation of strikers and nonstrikers. Unpublished paper, Patna University, 1969.

KIKUCHI, A., & GORDON, L. V. Bureaucratic orientation of the Japanese adolescent. *Japanese Annals of Social Psychology*, 1970, **11**, 217–229.

MERTON, R. K. Bureaucratic structure and personality. *Social Forces*, 1940, **18**, 560–568.

ORWELL, G. *Nineteen eighty-four*. New York: Harcourt Brace, 1949.

ROKEACH, M. *The open and closed mind*. New York: Basic Books, 1960.

WEBER, M. *Essays in sociology*. (Trans. by H. H. Gerth & C. Wright Mills) New York: Oxford University Press, 1946.

(Received December 17, 1969)

---

ment and positive correlations with favorableness of attitudes toward school were consistently obtained.

Journal of Educational Psychology
1971, Vol 62, No. 1, 67–70

INFLUENCE OF KNOWLEDGE AND TYPE OF OBJECTIVES ON SUBJECT-MATTER LEARNING

JOSEPH R. JENKINS
University of Delaware

AND

STANLEY L. DENO[1]
University of Minnesota

To determine what effects type and knowledge of the objectives of instruction have on learning two experimental variables (type and knowledge of objectives) were combined in a 2 × 3 factorial arrangement of treatments. One hundred and twelve subjects were randomly assigned to one of six experimental conditions or a learning control condition. The objectives used were either general or specific (type), and were provided either to teachers only, to teachers and students, or to students only (knowledge). Although significant learning occurred for all instructed conditions compared to a test-only control, neither knowledge of the objectives nor type of objectives differentially influenced performance on a criterion test.

A great deal has been written on the importance of stating educational goals in terms of observable student performance. Among the arguments for using behaviorally stated objectives are (*a*) unless a teacher (or seahorse; Mager, 1962) knows exactly where he is going he may find himself somewhere else; (*b*) behaviorally stated objectives reduce wasted time in temporary diversions, ephemeral entertainment, or other irrelevancies (Popham, 1968); and (*c*) stating goals in behavioral terms permits a more objective measurement of goal attainment.

Implicit in most of the recommendations for using behavioral objectives is the assumptions that with such objectives instruction will be improved by decreasing the time required for learning a particular subject matter, or by increasing the amount of subject matter acquired under conditions of fixed study time. Presently, however, more verbiage than data has spilled on this assumption and the general topic of behavioral objectives. The logical arguments for using behavioral objectives, which are compelling, would be enhanced with some empirical data.

The present study was designed to determine the empirical validity of the assumption that providing either teachers, students, or teachers and students with general or specific instructional objectives increases the amount learned during fixed instruction time. If knowledge of and type of objectives are significant instructional variables, then as students' and teachers' knowledge of specific objectives increases so should learning increase.

METHOD

Design

Two experimental variables were arranged in a 2 × 3 factorial design. The first variable, type of objective, consisted of (1) general (nonbehavioral) and (2) specific (behavioral) objectives. The second variable, knowledge of objective, consisted of (1) teacher knowledge (i.e., possession) of objectives prior to and during instruction, (2) teacher knowledge (as in Level 1) and student knowledge (i.e., possession) during instruction, and (3) student knowledge alone during instruction. It is well to note here that Level 3 (student knowledge alone) is confounded with the manner in which content was presented. In the third level, teachers were eliminated, and on the day of instruction students received both the objectives and a set of teaching materials from which to study. The teaching materials provided the students were identical to the materials given in advance to the teachers in the other conditions.

A seventh group, a learning control condition, was made up of subjects who were neither taught by teachers nor given study materials, but who took the criterion examination.

Subjects

The 112 students who served as subjects were volunteers from a sophomore level educational psychology course. For their participation these students received points toward their final grade. In addition, 16 secondary education students from a senior level educational psychology course served as teachers. The teachers, by participating in the experiment, also earned points for their final grade. The teachers were seniors who had completed most of their course work except student teaching. They were familiar with the content of the instructional materials from previous course work in education and psychology.

---

[1] The authors wish to express their gratitude to Margaret Fetzer and Peter Idstein for their assistance in the data collection.

Requests for reprints should be sent to Stanley L. Deno, Special Education Department, 109 Pattee Hall, University of Minnesota, Minneapolis, Minnesota 55455.

Materials

The objectives, the instructional materials, and the criterion examination were taken from an instructional unit developed by Baker (1967). Examples of general objectives are:

> This unit seeks to develop in students certain abilities which will make it easier for them to understand how knowledge in the social sciences is systematically accumulated. Broadly, the unit will help students to:
>
> 1. analyze written material critically
> 2. become familiar with some of the general procedures used in social science research
> 3. learn what common pitfalls in social science research must be avoided
> 4. know the prominent social scientists and their contribution.

The specific objectives were stated more precisely and each objective was accompanied by a sample test item. Some examples are:

1. From a set of alternative orders, the student should be able to select the proper ordering of the components of an experiment. These components are:

   Hypothesis
   Procedure
   Selection of Population and Sample
   Description of Sample
   Data Collection
   Data Analysis
   Statement of Results
   Interpretation of Results

Sample Test Item: DIRECTIONS: CHOOSE THE BEST ORDER OF STEPS FOR CONDUCTING AN EXPERIMENT.

a. analyze data, state results, form hypothesis
b. state results, analyze data, form hypothesis
c. form hypothesis, analyze data, state results
d. analyze data, form hypothesis, state results

2. Given the name of one of the data collection procedures listed below, the student should be able to select from alternatives the major advantage or limitation of the technique.

   Observations
   Interviews
   Questionnaires
   Projective Tests
   Statistical Records
   Mass Communication Records
   Experiments

Sample Test Item: DIRECTIONS: CHOOSE THE BEST ANSWER FOR THE FOLLOWING QUESTION. ON YOUR ANSWER SHEET FILL IN THE SPACE UNDER COLUMNS a, b, c, or d NEXT TO THE NUMBER OF THE QUESTION YOU ARE ANSWERING.

One limitation of using projective tests is:

a. the subjects easily find out the purpose of the test and do not behave naturally
b. the subjects may not make the same responses on repeated tests
c. the tests cannot be given to young children or other subjects who cannot read or write.

3. Given a description of the key contribution of a social scientist, the student should be able to identify this social scientist from sets of four alternatives which include:

   Thomas Carlyle
   Auguste Comte
   Sigmund Freud
   Herodotus
   Alexander von Humboldt
   John Locke
   Bronislaw Malinowski
   Karl Marx
   Margaret Mead
   Plato
   Ivan Pavlov
   Adam Smith

Sample Test Item: DIRECTIONS: CHOOSE THE SOCIAL SCIENTIST WHO FITS THE FOLLOWING DESCRIPTION.

> This historian produced the frontier theory of history. He suggested that the ever-expanding frontier with its limitless free land in America could be an explanation for the differences between Americans and Europeans.
>
> a. F. J. Turner
> b. Thomas Carlyle
> c. Bronislaw Malinowski
> d. Alexander von Humboldt

The test was modified from the original by the addition of several test items which were generated out of the original objectives and the deletion of other items. In all, the test contained 56 items. The test was not available to any of the participants until after instruction was completed.

Procedure

The teachers were individually contacted 2 weeks prior to the experiment and given the teaching materials. Each teacher was randomly assigned to one of four treatment combinations, general or specific objectives, and teacher or teacher-and-student knowledge of the objective, so that there were four teachers in each of the four conditions requiring teachers. When the teachers received their advance instructions they were provided with objectives, and in the appropriate conditions, with additional copies of the objectives that they would give to their students on the day of instruction.

Sixteen subjects were randomly assigned to each of the six treatment combinations and to the control. A teacher and four students were then assigned at random to one of 16 classrooms. In the two conditions where no teachers were present subjects were provided with objectives and instructional materials, and assigned to one of two classrooms where they could study. In the control conditions subjects were assigned to a classroom where they took the criterion test. In these later three conditions a monitor was present while the subjects studied and were tested.

In all conditions, except the control, subjects either studied or were taught for 1¾ hours, after which they were administered the criterion tests. All treatment occurred simultaneously. Subjects had unlimited time to complete the criterion test although no subjects took longer than 40 minutes.

Results and Discussion

The means and standard deviation for the number correct on the test for the six treatment conditions and the control condition are presented in Table 1.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE NUMBER CORRECT ON THE CRITERION TEST FOR GROUPS DIFFERING IN KNOWLEDGE OF THE OBJECTIVES AND THE TYPE OF OBJECTIVES

Type of objective	Knowledge of objectives					
	Teachers only		Teachers and students		Students only (no teachers)	
	M	*SD*	*M*	*SD*	*M*	*SD*
General	44.5	3.7	42.9	2.9	46.2	3.0
Specific	42.5	5.1	43.6	4.4	43.9	4.2
Control	38.6	5.5				

To ensure that learning had occurred performance of all experimental subjects was compared against the performance of learning control subjects (test only). The mean score for the experimental subjects reliably exceeded that for control subjects on the criterion test ($F = 21.2$, $df = 1/105$, $p < .001$).

A 2 × 3 analysis of variance was conducted to test the significance of the main effects—type of objective and knowledge of objectives—and their interaction. Neither the main effects nor interaction were statistically significant (type of objective, $F = 2.1$, $df = 1/90$, $p > .10$; knowledge of objectives, $F = 1.90$, $df = 2/90$, $p > .10$; interaction, $F = 1.3$, $df = 2.90$, $p > .10$).

The results clearly do not support the proposition that type and knowledge of the objectives of instruction facilitate learning. It is also interesting to note that students learned at least as much when they studied without the aid of a teacher. The nonsignificant findings cannot be explained away by suggesting that subjects were unmotivated and therefore were inattentive to instruction since significant learning did occur. Further, since the mean number of errors committed by the experimental groups ranged from 9.8–13.5, the nonsignificant differences are probably not the result of a ceiling effect.

The possibility remains, however, that type and knowledge of objectives were insignificant variables because they received inadequate attention from both the teachers and the students. Since teachers and students rarely are exposed to the explicit objectives of instruction they might fail to use these objectives appropriately either because their value is not recognized or because one must learn how to use explicit objectives.

Perhaps a stronger test of the benefits potentially derivable from explicit objectives could be obtained if students or teachers received some incentive to use the objectives, or were given practice in their use.

At least one other possibility exists for why knowledge and type of objective failed to emerge as significant variable. When a unit is well structured, that is, designed to facilitate the attainment of particular objectives, explicitly stated objectives may be superfluous in that teachers and students are able to "read through" the materials to the objectives. If this is the case specific objectives will influence learning only indirectly through their influence on the design of the curricular materials.

An additional conclusion suggested by the data is that the argument which suggests that explicitly stating behavioral objectives produces improvement in learning is a difficult argument to support empirically.

REFERENCES

BAKER, E. L. *Social science research methods: An experimental instructional unit.* Los Angeles: University of California, Department of Education, 1967.

MAGER, R. F. *Preparing instructional objectives.* Palo Alto: Fearon Publishers, 1962.

POPHAM, W. J. Probing the validity of arguments against behavioral goals. Paper presented at the annual convention of the American Educational Research Association, Chicago, February 1968.

(Received December 17, 1969)

Journal of Educational Psychology
1971, Vol 62, No. 1, 71–78

IDENTIFICATION, ORGANIZATION, AND READING COMPREHENSION FOR GOOD AND POOR READERS[1]

ROBERT OAKAN, MORTON WIENER[2]
Clark University
AND WARD CROMER
Wellesley College

Starting with a question about the relationship of word-identification skill, organization, and reading comprehension, the comprehension of good and poor readers in the fifth grade was compared for material presented visually and auditorially and under conditions where the subject had material paralleling the "input" of good and poor readers. Results failed to support the assumption that good identification is a sufficient condition for good comprehension and suggested that a significant amount of the comprehension difficulties of the poor readers may be attributable to the manner in which the input is organized. Implications for understanding the complex behaviors labeled reading are discussed.

All too often in the literature on reading, much emphasis is placed on the mastery of word identification[3] skills, while the importance of the organizational aspects involved in reading comprehension appears to be overlooked. All too frequently approaches to the teaching of reading, as well as to reading remediation, focus on specific skill deficiencies, particularly on the training of identification skills. In fact, many writers and theorists assume a direct relationship between identification and comprehension and even maintain that identification skills constitute the major antecedent to comprehension. For example, in a study of reading achievement among beginning readers, Durrell (1958) concludes that: "Most reading difficulties can be prevented by an instructional program which provides early instruction in letter names and sounds, followed by applied phonics and accompanied by suitable practice in meaningful sight vocabulary and aids to attentive silent reading [p. 5]." Gates (1947), while stressing that proficiency in comprehension comes through learning to read in terms of "thought units" (i.e., phrases), goes on to say that: "The ability to read by thought units comes as a natural result of gradually increasing efficiency in recognizing single words.... The deficiency in recognition of thought units grows out of the inadequate development of the techniques of recognizing isolated words and of using context clues (to aid in their recognition) [pp. 335–336]."

While the points raised by Gates and by Durell may be applicable to the acquisition stage of reading, many workers appear to assume that comprehension difficulty at *any* level of reading proficiency is attributable to identification skill deficiency. Although it is clear from the above quotations that investigators recognize the role played by additional activities in reading, the methods typically advocated by them stress heavily the mastery of word identification or "code-breaking" skills, assuming that once these skills are mastered, good reading comprehension will automatically follow.

---

[1] This investigation was supported in part by Grant Number OEG-1-9-080043-0005 (010), Project Number 8-A-043, from the Office of Education, Department of Health, Education, and Welfare. The authors are indebted to Mary McAuliffe, Arthur Sullivan, Eleanor Moosey, and the principals, teachers, and pupils in the Worcester Public School System who made this study possible.

[2] Requests for reprints should be sent to Morton Wiener, Department of Psychology, Clark University, Worcester, Massachusetts 01610.

[3] "Identification" in reading may be defined as the process involving the ability to discriminate among graphic symbols, among auditory symbols (i.e., among words presented both visually and auditorily), and the ability to transform such symbols from a visual to an auditory form. If the products of such a process of identification correspond to one's available auditory language forms, meaning can then be associated with the visual forms. (See Wiener & Cromer, 1967.)

In contrast to the view which emphasizes a "natural" identification-comprehension relationship, several writers (e.g., Fries, 1963; Goodman, 1964; Lefevre, 1964; Wiener & Cromer, 1967), while recognizing the contribution of a proficiency in identification for the beginning reader, have given more emphasis to the organizational aspects of language which they consider necessary for the achievement of good reading comprehension. As Wiener and Cromer note, some difficulties in reading comprehension can be accounted for in terms of a mismatch between the reader's typical patterns of linguistic organization and the organizational patterns required for the comprehension of the particular written material. It is held that this kind of poor reader has developed language patterns such that his organization and elaboration of the printed language is different from that of the good reader, either because he has not learned a consensual response pattern, or because he has learned an idiosyncratic pattern too well. This formulation implies that either (*a*) a change in the way in which written material is organized, or (*b*) a change in one's response patterns to such material will result in improved reading comprehension. If organization is critical, then the attainment of a high level of identification skills does not necessarily imply a similarly high level of reading comprehension.

Cromer (1968) delineated two groups of readers who performed relatively poorly in terms of reading comprehension. One of these groups had a vocabulary skill deficit; the second was made up of readers who demonstrated adequate vocabulary skills but who were hypothesized to have organizational difficulties. The readers in the second group were assumed to organize their "input" in a linguistically nonrelevant way, that is, they read in a word-by-word fashion rather than organizing input into such "meaningful units" as phrases, and consequently they lose that part of the meaning carried by combinations of words.[4]

One of Cromer's hypotheses was that if reading material were preorganized into meaningful units (i.e., if the graphic forms were arranged in phrases), the level of comprehension of these poor readers would be increased relative to their understanding of reading passages presented in the usual format. As hypothesized, comprehension for this group of poor readers, but not for the vocabulary deficit group, was significantly improved under the preorganized condition. Moreover, with this type of input this group did as well as did a matched group of good readers under any input condition. Cromer also explored his data for the relationship of identification to comprehension and found for these readers, at least, no support of the view that good reading comprehension is a direct function of good identification.

An earlier study by Cromer and Wiener (1966) further suggests that a high level of identification skills may not be a sufficient condition for the occurrence of adequate reading comprehension. In one part of their study, Cromer and Wiener utilized a Cloze technique (Taylor, 1953) whereby words were removed randomly from printed stories. The subjects (matched good and poor readers in the fifth grade) were asked to read these passages aloud and give an appropriate word to fit each blank. The experimenter provided the subject with words he could not identify, corrected any reading errors which were important in the content of the story, and offered an appropriate word whenever a subject indicated he could not supply one for a particular blank. Analysis of the data showed that the good readers as a group gave a significantly greater number of consensually and syntactically correct word insertions than did the poor readers. Insofar as a high degree of accuracy of identification was assured for each good and poor reader alike in the study, and the subjects were guided in their insertions of words by their general comprehension of the stories, this finding is contrary to what would be expected if there were a direct relationship between identification and comprehension for all readers.

In terms of information processing, one's comprehension of a communication, whether in graphic or auditory form, can be considered in terms of the adequacy of the "input" of the message. For purposes of the present analysis, such input may be considered to have two distinct aspects: the "identification" of the words comprising the message and the "organization" of these words into patterns or units. By criteria of consensuality within one's language, both identification and organization in the processing of a communication may be said to be either "good" or "poor." For example, readers who have no difficulty in comprehending written material may be assumed to have "good identification" and "good organization," both of which contribute to "good visual input" when reading. On the other hand, readers showing reading comprehension problems may demonstrate difficulties in identification, in organization, or in both. By this analysis, relatively poor comprehension may occur for a given reader when there is (*a*) poor identification and poor organization, (*b*) good identification and poor organization, or (*c*) poor identification and good organization.

---

[4] In both written and spoken language, meaning is not carried solely by the individual lexical units. A highly significant amount of meaning is delineated by the patterning of individual words into larger structural units.

Thus far the focus in most investigations has been on input within the visual mode. However, comprehension of communications given in the auditory mode also may be analyzed in the above terms. For example, when a story is read aloud to someone, the listener's comprehension is not only a function of his own skills, but also a function of the speaker's skill in identification *and* in organization. Thus, even though the listener may be an accomplished identifier and organizer himself, he may have some difficulty in comprehending material read aloud by a person who demonstrates (*a*) poor identification and poor organization, (*b*) good identification and poor organization, or (*c*) poor identification and good organization. In each of these cases "poor auditory input" may be said to occur for the listener.

The present study was designed to assess the relationship of identification and organization to comprehension for both good and poor readers. Both level of identification and patterns of organization were systematically varied to assess their relation to comprehension in both the visual and auditory modes. Four questions were investigated:

1. Although the present authors maintain that a high level of word identification skill is not sufficient for the occurrence of good comprehension for all readers, many workers in the field of reading make, at least tacitly, just such an assumption. The hypothesis that identification training leads to an increase in the comprehension of poor readers can be assessed.

2. While it is commonly assumed that a high level of identification is necessary for good reading comprehension for all readers, it is held that word identification may be a relatively unimportant aspect of reading comprehension for the more accomplished reader (Wiener & Cromer, 1967). It is hypothesized, then, that if a high level of identification is necessary for the comprehension of good readers, having good readers read under conditions of identification similar to that of poor readers lessens the level of comprehension for these subjects.

By manipulating the level of visual identification while holding constant the organizational patterns of good and poor readers, these aspects of the identification-comprehension relationship can be evaluated.

3. It can be assumed that good and poor readers use differential patterns in organizing input and, if so, then this difference should be manifest in their handling of material presented auditorily as well as visually. It is hypothesized that the comprehension of both good and poor readers is impaired under conditions of poorly organized auditory input, but that poor readers show a differentially greater degree of impairment under these circumstances.

4. The possibility remains that the difficulty in reading comprehension encountered by the poor reader can be attributed to some deficit or defect which interferes with his comprehension in general, that is, a condition independent of his attainment of good identification skills and of his patterns of linguistic organization. If this possibility is tenable, and poor readers have some general comprehension difficulty, then comprehension by these individuals of auditorily presented material in which identification and organization is optimal should also be

lower than that of good readers under similar good input conditions.

METHOD

Subjects

Fifty-four boys and 42 girls in the fifth grade in a New England public school system served as subjects in the study. Half the subjects were poor readers, reading (i.e., comprehending) on a level at least 1½ years below grade level, as measured by standardized reading tests administered by the schools. The second group of subjects were reading at or above grade level, and showed no reading comprehension difficulties. A matched-pairs technique was used, with each pair of good and poor readers coming from the same class in school, being of the same sex, and of approximately the same age and intelligence. Half of the good and poor readers participated in Part 1 of the study, the other half in Part 2. Reading level scores for the poor readers ranged from 2.5 to 3.5 with a mean of 3.25. The reading level for the good readers ranged from 5.0 to 8.8, with a mean of 6.1. Both groups of subjects ranged in age from 9 years 10 months to 11 years 3 months, with a mean age for each group of 10 years 6 months. The intelligence test scores for both groups of subjects, on a fourth-grade Otis test, varied from 90 to 110, with means of 99 and 98 for the good and poor reader groups, respectively.

General Procedure

Each subject was tested within his own school building during school hours. Part 1 of the study required approximately 20 minutes for each of the good readers; from 30 minutes to an hour was required for each of the poor readers, depending upon the amount of identification training necessary. Part 2 of the study required approximately 25 minutes for both the good and poor readers. The order of all presentations was fully counterbalanced.

Efforts were made to maintain an informal situation and to put each of the children at ease. Each child was told that what he was to do had no bearing on his grades or promotion. No child appeared to become upset by any aspect of the experimental procedure, nor was there any evidence that any child was not performing at his best.

Materials and Procedure

The four reading passages used in the study were adapted from standardized reading tests appropriate to the fifth-grade level. Comprehension of these passages was measured by five multiple-choice type questions for each story.

Part 1 of the study was designed to assess the effects on reading comprehension of (*a*) improved identification for poor readers; and (*b*) impaired identification for good readers.

Each poor reader in Part 1 was given a set of stories to read, presented in the regular prose form, and their comprehension of this material was tested; this provided a base-line measure of reading comprehension for these subjects. As it is generally held that poor readers are deficient in word identification skills, it may be said that these children have poor visual input (at least in terms of identification) when reading. This group of poor readers was then trained to identify each individual word occurring in a second set of stories which they were subsequently asked to read. It may be said that with such identification training, good visual input (with respect to identification) was provided for the subjects under these circumstances.

For use in the above identification training, four sets of flash cards were made, one for each of the stories. The words appearing in the reading passages were individually typed on these cards. Each poor reader in Part 1 was trained, using the flash cards, so that he could correctly identify (read aloud) each of the words occurring in the stories he would later be asked to read. The criterion for a subject's having learned the words was a minimum of three successively correct identifications of each word. The more difficult of these words were briefly defined and used in a simple sentence.

As with the poor readers above, the good readers in Part 1 of the study were given a set of stories to read, and a base-level of comprehension was established for these subjects. As it is generally held that good readers have adequate identification skills, it is assumed that these subjects have good visual input (with respect to identification) when reading such material. The comprehension of these subjects was then tested for a set of passages designed to provide poor visual input in terms of identification. This set of stories was constructed by first tape recording several poor readers in the fifth grade reading the four passages aloud. Typed transcripts of these tapes were then made, including all of the poor reader's pauses, false starts, errors, mispronunciations, omissions, etc.

Part 2 of the study was designed to investigate the relationship of identification and organization to comprehension for material presented in the auditory mode.

The comprehension of both the good and poor readers in Part 2 was tested for tape-recorded passages in which organization was poor. This poor-auditory-input material was recorded by two male and two female poor readers in the fifth grade. Their readings of the four stories were replete with errors: mispronunciations, pauses and false starts, the adding and omission of words. However, each time an error was made, the experimenter corrected the reader, saying the appropriate word(s) aloud. Thus, the proper identification of each word in the stories was made, while the poor readers' generally disrupted and disorganized style of reading remained intact.

Each good and poor reader in Part 2 also listened to tape recordings of good readers (i.e., readers good in both identification and organization) reading a set of stories and answered questions measuring their comprehension of these passages. These good-auditory-input recordings were made by two male and two female readers in the fifth grade. Their readings of the four passages were generally quite accurate; the few errors which were made were all self-corrected, and were judged to be insufficient to impair one's comprehension of the material in any way.

Each pair of subjects either read or listened to two of the four passages under the good input condition, and the remaining two stories under the poor input condition. As there were five questions for each story, a subject could achieve a maximum comprehension score of 10 on each good and poor input pair of stories. The quantitative data for the study consisted of the number of questions answered correctly on each of the two pairs of stories presented to the subjects.

RESULTS

To test the difference between the experimental conditions (auditory versus visual modes of presentation, and good versus poor types of input), an analysis of variance using a three-factor mixed design with repeated measures on one factor (Bruning & Kintz, 1968) was employed. A summary of the results of the analysis is presented in Table 1. The good readers answered a significantly greater ($F = 30.16$, $df = 1/92$, $p < .001$) number of questions correctly than did the poor readers. The effect of mode of presentation (i.e., auditory versus visual) was not significant ($F = 2.42$, $df = 1/92$, $p > .05$). However, the interaction of the good and poor reader groups with the visual and auditory mode of presentation (Group × Mode interaction) was significant ($F = 6.61$, $df = 1/92$, $p < .025$). The good readers answered more questions correctly on the stories presented visually ($M = 6.46$) than on those presented auditorily ($M = 5.50$), while no such difference was evident for the poor readers (visual: $M = 4.33$; auditory: $M = 4.83$). The effect of type of input (i.e., good versus poor input) was found to be significant ($F = 9.26$, $df = 1/92$, $p < .005$). Neither the interaction of the good and poor reader groups with the good and poor types of input (Group × Input interaction), nor the interaction of the visual and auditory modes of presentation with the good and poor types of input (Mode × Input interaction) was significant (for both interactions, $F = < 1$, $df = 1/92$). However, the interaction of the good and poor reader groups, the visual and auditory models of presentation, and the good and poor types of input (Group × Mode × Input interaction) was significant ($F = 15.98$, $df = 1/92$, $p < .001$). The means for these groups are given in Table 2. An examination of these means shows that the good readers demonstrated their highest level of comprehension under the condition of good visual input, while the poor readers did best under the good-auditory-input condition. Identification training (good visual input) did not result in any apparent increase in comprehension for the poor readers. The good readers, however, did perform relatively less well on the poor-visual-input material. While the good and poor readers demonstrated a comparable level of comprehension on the good-auditory-input passages, the performance of the poor readers fell under the poor auditory condition, in

TABLE 1

ANALYSIS OF VARIANCE OF NUMBER OF QUESTIONS CORRECTLY ANSWERED BY GOOD AND POOR READERS UNDER CONDITIONS OF AUDITORY VERSUS VISUAL MODES OF PRESENTATION AND GOOD VERSUS POOR TYPES OF INPUT

Source	*df*	*MS*	*F*
Between subjects	95		
Group (G)	1	93.5	30.16***
Mode of Presentation (M)	1	7.5	2.42 (*ns*)
G × M	1	20.5	6.61*
Error	92	3.1	
Within subjects	96		
Type of input (I)	1	17.5	9.26**
G × I	1	0.1	1 (*ns*)
M × I	1	0.0	1 (*ns*)
G × M × I	1	30.2	15.98***
Error	92	1.89	
Total	191		

* $p < .025$.
** $p < .005$.
*** $p < .001$.

TABLE 2

MEAN NUMBER OF QUESTIONS ANSWERED CORRECTLY, GROUP × MODE × INPUT INTERACTION

Poor readers				Good readers			
Auditory		Visual		Auditory		Visual	
Good	Poor	Good	Poor	Good	Poor	Good	Poor
5.33	4.33	4.42	4.25	5.29	5.71	7.29	5.63

contrast to that of the good readers who showed no such trend.

DISCUSSION

The findings of the study lend support to the argument that an appreciable amount of the poor reader's comprehension difficulties may be attributable to the manner in which he organizes his input. No support was found for the assumption of a direct relationship between identification and comprehension for any group of readers.

The hypothesis that if reading comprehension is a direct function of identification, then training poor readers to identify all the words found in stories they are subsequently asked to read should lead to a higher level of comprehension for this material (relative to their comprehension of similar passages not prefaced by such training) was tested. No appreciable improvement in comprehension was found for the poor readers who received such identification training. Thus, it was not demonstrated that good identification is a sufficient condition for good comprehension for all readers.

The hypothesis that under conditions of impaired identification the comprehension of good readers would be less than that for similar material presented under normal conditions was supported. The level of comprehension of the good readers fell off markedly under the poor-visual-input condition.

Thus, while the results from Part 1 of the study tend to confirm the observation that poor identification (at least when confounded with poor organization) is associated with poor comprehension, the contention that good identification is invariably related to good comprehension was not substantiated.

Only partial support was found for the hypothesis that the comprehension of both good and poor readers is impaired under conditions of poor auditory input (poorly organized material). While the poor readers did show a lower level of comprehension for this condition, relative to good auditory input, the level of comprehension of the good readers did not decrease. Thus, it would appear that the comprehension of the poor readers suffered under conditions of poorly organized auditory input, while that of the good readers was not affected to any significant degree.

There was no support for the hypothesis that if poor readers have some general deficit in comprehension, then their comprehension of material presented auditorily is lower than that of good readers for the same material. Good and poor readers appear to perform equally well on the good-auditory-input material.

The findings of the study may be interpreted within one framework of reading difficulty suggested by Wiener and Cromer (1967) which also includes a linguistic analysis of reading.

According to Wiener and Cromer's Difference Model, problems in reading comprehension may be a product of the manner in which the reader organizes his input: there is a mismatch between one's typical modes of responding to written material and the patterns of responding assumed necessary for adequate comprehension to occur. In terms of this model, therefore, a reader who is skilled in word identification may still exhibit comprehension difficulties if he does not organize his reading input into certain patterns.

Considerable support is given to the Difference Model by modern linguistic analyses of reading. Workers such as Fries (1963) and Lefevre (1964) point out that while the majority of current approaches to reading methodology focus on the word as the most significant meaning-bearing unit, linguistically, the word is considered a relatively minor language unit. Lefevre (1964) states that many poor readers miss patterns of meaning in the material they read because they miss the meaning-bearing patterns of language:

> If they do not literally read word-by-word, they often read by arbitrary word groups or sentence fragments that make almost as little sense as isolated words called out one at a time.... All these structural errors cause failures in comprehension, since meaning cannot be reached except through the structures that carry meaning [p. 23].

Both Lefevre and Fries maintain that patterns of intonation constitute one of the most significant systems by which meaning is signaled in language. Often these patterns of pitch, stress, and pause furnish the minimum contrasts that identify and separate the particular meanings of a communication. Fries (1963) relates this to the reading process as follows:

> On the whole, contrary to the beliefs of many, written material contains less of the language signals than does talk. In the graphic representations of language there are left out such language signals as intonation and stress and pause.... If one is to read with comprehension the graphic representations of the language signals, he must learn to supply these portions of the signals which are not in the graphic representations themselves [p. 130].

The assertions of the Difference Model and of the linguistic analysis of reading provide a means by which many aspects of the present study may be understood.

It was found in Part 1 of the study that for poor readers, identification training was not sufficient to improve reading comprehension. Following a linguistic analysis, it can be noted that printed material is preorganized only with respect to features of punctuation—it is entirely up to the reader to supply all other aspects of organization which mark and signal the particular meanings carried by the communication. Thus, even if they receive identification training, if poor readers typically do not organize their input into certain efficacious patterns, they may have considerable difficulty understanding what they read—good identification is then not directly related to good comprehension under such circumstances.

A similar interpretation may be made for the findings in Part 2 of the study. When listening to material in which the words were both well-identified and well-organized, poor readers demonstrated a level of comprehension that was fully comparable to that of good readers for the same material. In this case, even if they are assumed to be generally deficient in their own abilities to organize, or if they typically organize "differently," the high level of organization imposed on the material by the readers of the passages may have proved sufficient to guide the comprehension of the poor readers. When listening to the poor-auditory-input stories, in which organization was poor, the good readers showed as high a level of comprehension as they did for the good-auditory-input material, while the comprehension of the poor readers markedly decreased. This finding seems to imply that the good readers impose order on their input; the response patterns of the poor readers, however, did not appear to be sufficient to compensate for the lack of proper organization of the material.

One unexpected finding in the study was that the good readers demonstrated a higher level of comprehension for the good-visual-input material ($M = 7.29$) than for the good-auditory-input passages ($M = 5.29$). If Fries is correct in stating that written material contains fewer language signals than does talk, one would have expected the good readers to have shown their highest level of comprehension on the good-auditory-input passages. One possible explanation for this is that the subjects had differing amounts of exposure time to the visually and auditorily presented stories. When reading the passages themselves, the subjects may have been able to go over words and sentences as many times as they required; when listening to the taped stories, however, they could only hear a single presentation of the material.

Perhaps the most general implication of

the present study is related to the teaching of reading. While it was seen that poor identification could be a precursor to poor comprehension, there was no evidence in the study which would support the contention that good identification is systematically related to good comprehension for all readers. This would suggest that for a certain class of reader, methods of instruction which primarily emphasize word identification skills are not sufficient for the development of a high level of reading comprehension. In fact, this focus upon the word may impede the imposition of organization even after identification skill is mastered. Thus, any method of reading instruction which first emphasizes word-by-word identifications, and only later, if at all, comprehension (as do many contemporary approaches) may itself discourage the process of organizing input into the meaningful units which guide one in understanding what has been read.

In a recent study, Cromer (1968) demonstrated that readers who have not adequately learned to deal with written material in terms of meaningful patterns or organization can be encouraged to do so, without extensive training, by the simple method of grouping the material for them. Although Cromer's study involved relatively accomplished readers, this technique of preorganizing reading materials may also prove to be beneficial for the beginning reader as well. A more thorough approach to the teaching of reading might involve, in addition to the training of identification skills, the use of books in which the contents were printed in preorganized word groupings, and listening sessions in which the organizational units of good readers were emphasized. Above all, a child should be taught that words derive part of their meaning within the patterns and groupings in which they occur, and instruction as to the means of attending to such organizational cues has to be included as an integral part of and co-jointly with the teaching of all reading skills.

REFERENCES

Bruning, J. L., & Kintz, B. L. *Computational handbook of statistics*. Glenview, Ill.: Scott, Foresman, 1968.

Cromer, W. The effects of "pre-organizing" reading material on two types of poor readers. Unpublished doctoral dissertation, Clark University, 1968.

Cromer, W., & Wiener, M. Idiosyncratic response patterns among good and poor readers. *Journal of Consulting Psychology,* 1966, **30,** 1–10.

Durrell, D. D. First-grade reading success study: A summary. *Journal of Education,* 1958, **140,** 2–6.

Fries, C. C. *Linguistics and reading*. New York: Holt, Rinehart & Winston, 1963.

Gates, A. I. *The improvement of reading*. New York: Macmillan, 1947.

Goodman, K. S. The linguistics of reading. *Elementary School Journal,* 1964, LXIV, 355–361.

Lefevre, C. A. *Linguistics and the teaching of reading*. New York: McGraw-Hill, 1964.

Taylor, W. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly,* 1953, **30,** 415–433.

Wiener, M., & Cromer, W. Reading and reading difficulty: A conceptual analysis. *Harvard educational Review,* 1967, **37,** 620–643.

(Received December 27, 1969)

Journal of Educational Psychology
1971, Vol 62, No. 1, 79–80

WEIGHTED AND UNWEIGHTED MEANS FOR ESTIMATION:

A NOTE ON THE HUMPHREYS-DACHLER AND JENSEN PAPERS[1]

THOMAS J. FISCHBACH[2] AND HERBERT J. WALBERG
University of Wisconsin

The methods used by Humphreys and Dachler to estimate effects in their analysis of Project TALENT data to test Jensen's Theory of Intelligence are shown to produce biased estimates. A statistical model appropriate for the problem and Humphreys and Dachler's definitions of parameters is presented and alternative estimators which are unbiased and have other desirable statistical properties are proposed.

Humphreys and Dachler (1969) have taken issue with Jensen's (1968) use of a "pseudo-orthogonal" design and statistical analysis and have reported a weighted-means analysis of Project TALENT data on ninth-grade boys which, they claimed, refutes the Jensen theory of intelligence. While other investigators are now indebted to both Humphreys and Dachler and to Jensen for clarifying several important methodological issues which point to more definitive tests of the hypothesis, the results of the analysis by Humphreys and Dachler are subject to systematic errors which have not been noted in Jensen's (1969) reply. The following exposition is limited to the statistical analysis used by Humphreys and Dachler and is not addressed to the question of the relevance or lack of relevance of the data for tests of Jensen's theory or to other issues raised by either Humphreys and Dachler or Jensen.

Humphreys and Dachler (1969) advocate estimation of the effects of intelligence (IQ), socioeconomic status, and the interaction by linear combinations of cell means weighted by weights proportional to the cell frequencies rather than the usual linear combinations of the unweighted cell means. It is easy to show, however, that these estimators are biased while the ones based on the unweighted cell means are not.[3] Indeed, the latter are in fact least-squares estimators of the corresponding effects (see for example Scheffé, 1959, Corollary 1, p. 15) since the cell means are least-squares estimators of the true values for each cell.

To show that the Humphreys-Dachler estimators are biased one need only state an appropriate statistical model and substitute terms in the equations for their estimators. For example, one might parameterize the problem in terms of the following model:

$$Y_{HHk} = \mu + \tfrac{1}{2}\alpha^{IQ} + \tfrac{1}{2}\beta^{SES} + \tfrac{1}{4}\gamma + e_{HHk} \quad [1]$$

$$Y_{HLk} = \mu + \tfrac{1}{2}\alpha^{IQ} - \tfrac{1}{2}\beta^{SES} - \tfrac{1}{4}\gamma + e_{HLk} \quad [2]$$

$$Y_{LHk} = \mu - \tfrac{1}{2}\alpha^{IQ} + \tfrac{1}{2}\beta^{SES} - \tfrac{1}{4}\gamma + e_{LHk} \quad [3]$$

$$Y_{LLk} = \mu - \tfrac{1}{2}\alpha^{IQ} - \tfrac{1}{2}\beta^{SES} + \tfrac{1}{4}\gamma + e_{LLk} \quad [4]$$

where

Y_{HLk}, for example, refers to the test score of the *k*th individual of high IQ (the first subscript) and low socioeconomic status (the second subscript),

α^{IQ} is the effect of IQ,

β^{SES} is the effect of socioeconomic status,

---

[1] The research and development reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare, under the provisions of the Cooperative Research Program (Center No. C-03, Contract OE 5-10-154).

The authors wish to acknowledge helpful comments from R. Darrell Bock and Joel Levin. Of course, this does not imply their agreement with all points in the paper.

[2] Requests for reprints should be sent to Thomas Fischbach, R & D Center for Cognitive Learning, University of Wisconsin, Madison, Wisconsin 53706.

[3] This should not be confused with the so-called "method of unweighted means."

γ is the interaction effect, and the error terms are assumed to be independent, with zero means and equal variances (but not necessarily normally distributed).

To illustrate the bias in the Humphreys-Dachler estimators consider their estimator for the socioeconomic status effect which is termed $\beta^{*\mathrm{SES}}$ in this note. This is given by:

$$\beta^{*\mathrm{SES}} = [(n_{HH}\bar{Y}_{HH\cdot} + n_{LH}\bar{Y}_{LH\cdot})/n_{\cdot H}] - [(n_{HL}\bar{Y}_{HL\cdot} + n_{LL}\bar{Y}_{LL\cdot})/n_{\cdot L}] \quad [5]$$

where

$\bar{Y}_{HH\cdot}$ is the mean for the high-high cell with the other means similarly defined,

n_{HH} is the cell frequency for the high-high cell and the other cell frequencies are similarly defined, and

$n_{\cdot H}$ is the total frequency for high socioeconomic status.

By substitution of Equations 1, 2, 3, and 4 in Equation 5 it is not difficult to obtain

$$\beta^{*\mathrm{SES}} = \beta^{\mathrm{SES}} + \tfrac{1}{2}\alpha^{\mathrm{IQ}}\left(\frac{n_{HH} - n_{LH}}{n_{\cdot H}} + \frac{n_{LL} - n_{HL}}{n_{\cdot L}}\right) + \tfrac{1}{4}\gamma\left(\frac{n_{HH} - n_{LH}}{n_{\cdot H}} - \frac{n_{LL} - n_{HL}}{n_{\cdot L}}\right) + e^{*} \quad [6]$$

where e* is a linear combination of error terms.

The expectation of Equation 6 is not β^{SES} but contains bias terms in α^{IQ} and γ. Since IQ and socioeconomic status are positively correlated it is clear that the coefficient of α^{IQ} in Equation 6 is positive (in fact for the data given by Humphreys and Dachler the coefficient of α^{IQ} is + 0.3), so that their socioeconomic status effect is positively biased by a positive IQ effect. This explains the changes they reported. Simple algebra can be used to show similar biases for their estimates of the IQ effect and the interaction effect.

It can be shown that the least-squares estimators based on the unweighted cell means are not biased. For example,

$$\hat{\beta}^{\mathrm{SES}} = (\bar{Y}_{HH\cdot} + \bar{Y}_{LH\cdot} - \bar{Y}_{HL\cdot} - \bar{Y}_{LL\cdot})/2 \quad [7]$$

is the least-squares estimators of β^{SES} and

$$E(\hat{\beta}^{\mathrm{SES}}) = \beta^{\mathrm{SES}} \text{ by substitution from Equations 1, 2, 3, and 4.} \quad [8]$$ [4]

Thus, the average or expected value is not influenced by the size of any parameter except that for socioeconomic status. The same result holds for each of the other two least-squares estimators which Humphreys and Dachler rejected.

Humphreys and Dachler stress the importance of generalizations to general populations. The first step in doing this is the postulation of models such as those presented previously (along with certain simple regularity postulates to cover middle groups). Such models potentially describe the response of *each and every individual* in the population if efficient unbiased estimates of parameters are obtained. But it is clear that use of biased estimators can only result in fallacious generalizations to general populations regardless of the model.

Moreover, this model also implies that when designing a study there is no need to be concerned about the relative numbers of persons in each of the four populations. The best and most economical method of gaining information about the parameters is to sample each group equally (unless, of course, there are sampling cost differentials). Thus, the question of what the joint distribution of IQ and socioeconomic status is like becomes an irrelevant issue.

REFERENCES

HUMPHREYS, L. G., & DACHLER, P. Jensen's theory of intelligence. *Journal of Educational Psychology,* 1969, **60,** 419–426.

JENSEN, A. R. Jensen's theory of intelligence: A reply. *Journal of Educational Psychology,* 1969, **60,** 427–431.

JENSEN, A. R. Patterns of mental ability and socioeconomic status. *Proceedings of the National Academy of Sciences,* 1968, **60,** 1330–1337.

SCHEFFÉ, H. *The analysis of variance.* New York: Wiley, 1959.

(Received January 19, 1970)

---

[4] Also, it is not difficult to show that the square of the estimate divided by its standard error yields the optimal F test described by Scheffé (1959, pp. 116–119) for the unequal n case.

JOURNAL OF
EDUCATIONAL PSYCHOLOGY
MONOGRAPH

Vol. 62, No. 1 February, 1971

RETROACTIVE INHIBITION OF MEANINGFUL DISCOURSE[1]

RICHARD C. ANDERSON[2] AND DAVID L. MYROW
University of Illinois

Two experiments involving 297 high school students investigated retroactive inhibition with prose passages. In each case, retroactive inhibition was observed where consecutive passages entailed similar stimuli and competing responses while retroactive facilitation appeared where the passages involved both similar stimuli and similar responses. Previous failures to find retroactive inhibition with prose are probably attributable largely to the mixing of test items, some of which were facilitated, some of which were inhibited, and some of which were unaffected. The results matched expectations from interference theory fairly closely, except that multiple-choice items were more subject to retroactive inhibition than short-answer items, suggesting that response competition is a more important factor than response unavailability in retroactive inhibition with prose.

The geneticist experiments with fruit flies and bread mold because these organisms are easier to study than, say, oak trees, elephants, or people. For similar reasons, lists of nonsense syllables and discrete words have long comprised the chief experimental tasks of psychologists interested in memory. While research with rote tasks is perhaps a self-perpetuating tradition, most experimental psychologists undoubtedly operate on the faith that their research is generally applicable to memory for verbal-symbolic material. Whether this faith is well placed is open to serious question on the basis of data now available. One of the central phenomena of memory for rote verbal tasks is retroactive inhibition, the increase in forgetting that occurs when a person learns a second, similar list after having acquired a to-be-remembered list. Though many investigators have tried, it has been very difficult to demonstrate retroactive inhibition in memory for meaningful, connected discourse, such as pages from a textbook or stanzas from a poem. This paper contains an analysis of theoretical and methodological problems that may have prevented previous research from detecting retroactive inhibition with meaningful discourse and then reports two experiments based on this analysis.

Slamecka (1960, 1962) has completed a series of studies of retroactive inhibition with connected discourse in which he has found that recall (*a*) increases as a function of the number of trials of original learning, (*b*) decreases as a function of the number of trials of interpolated learning, (*c*) decreases as a function of the number of interpolated passages, and (*d*) decreases as a function of the similarity between the original learning and the interpolated learning passages. Slamecka's results match closely the results obtained with paired associates

---

[1] Preparation of this paper was supported in part by the Advanced Research Projects Agency through the Office of Naval Research under Contract ONR Nonr 3985(08). The first experiment was reported in the junior author's bachelor's thesis, supported in part by the National Science Foundation under Grants GY2850 and GY4471. The authors are indebted to Thomas Andre, who assisted in the analysis of the data, and to Raymond Kulhavy, Sylvia Crooks, and Laura Lieberman, who helped collect the data. The authors are grateful for the cooperation of the staffs at Lakeview High School, Decatur, Illinois and Rich East High School, Park Forest, Illinois.

[2] Requests for reprints should be sent to Richard C. Anderson, Training Research Laboratory, University of Illinois, 8 Lincoln Hall, Urbana, Illinois 61801.

(cf. Keppel, 1968), but it can be argued that this is because his materials and procedures forced rote learning and recall. Below is an example of one of Slamecka's (1960) passages:

> We must postulate that from strictly semantic points of vantage, most confusions in communication revolve about inadequate stipulation of meaning [p. 104].

Passages such as this were learned and recalled one word at a time using the serial anticipation method, a procedure that could conceal the relationships among the words in already obscure material.

If Slamecka's experiments are discounted, the remaining evidence for retroactive inhibition with connected discourse is equivocal. Ausubel, Robbins, and Blake (1957) presented a 1,700-word passage dealing with Buddhist concepts of God, immortality, soul, faith, salvation, morality, and responsibility. Immediately thereafter a multiple-choice Buddhism achievement test was given. On the second day one group studied a 1,700-word passage on Christianity whereas another group received no interpolated material. The Buddhism test was repeated on the eighth day. The group that received the Christianity passage retained 82% whereas the control group retained 83%. Ausubel, Stager, and Gaite (1968) found that a passage judged to be "similar but conflicting" actually facilitated memory for a passage studied earlier. Subjects received a passage on Zen Buddhism, then either a passage on Buddhism or an irrelevant passage on drug addiction. After an interval of 7 days, a multiple-choice Zen Buddhism test was administered. Subjects who had received the Buddhism passage did significantly *better* on this test than those who had read the drug addiction passage. Gaite (1968) repeated this experiment three times. The nine groups in his experiments that received the Buddhism passage averaged .2% higher on the Zen Buddhism test than eight control groups. Several investigators have found retroactive inhibition with connected discourse; however, with one exception described below, the effect has been very small.

It would be premature to conclude as Ausubel and his associates have done (Ausubel, 1963; Ausubel et al., 1968) that the apparent failure to find retroactive inhibition indicates that the learning and remembering of connected discourse entails entirely different processes than the learning and remembering of rote tasks. There are several difficulties with the previous research. The biggest problem has been the failure to consider in detail the similarities and differences between the original learning and interpolated learning passages. Since Osgood's (1949) article, it has been recognized that retroactive inhibition occurs with paired associates when the stimuli in successive lists are identical or very similar and the responses are antagonistic. When both stimuli and response are very similar, there is retroactive facilitation. Little effect is observed when both the stimuli and responses in successive lists are different.

Stimulus and response similarity have not been analyzed in most of the studies of memory for meaningful discourse. If one thinks only of the original learning and interpolated learning passages, it is impossible to distinguish between stimuli and responses; however, the distinction is possible when the test is considered. The test-item stems define the stimuli. Stimulus and response similarity can be determined from content analyses of the passages in relation to the test items. The tests employed in the previous studies may have included some test questions which could be answered correctly on the basis of the interpolated learning passage and other questions to which the interpolated learning passage was irrelevant. Marked interference would be expected only on questions to which the interpolated learning passage leads to incorrect, competing responses.

Variability in the type of test procedure has been a second source of confusion in previous studies of retroactive inhibition with connected discourse. Recent research with paired-associate lists (cf. McGovern, 1964; Postman & Stark, 1969) seems to indicate that retroactive inhibition is primarily a matter of loss of availability of responses rather than specific associative interference. Recall procedures—for instance, short-answer or completion questions—are sensitive to response availability. Response availability is much less important when a recognition procedure, such as a multiple-choice test, is employed. Studies of memory for connected discourse which have employed a recall procedure have found at least some retroactive inhibition (Entwisle & Huggins, 1964; Hall, 1955; King & Cofer, 1960; McGeoch & McKinney, 1934; Mehler & Miller, 1964). No retroactive inhibition, or even retroactive facilitation, has appeared only when a multiple-choice test was used.

To parallel the research with paired-associate lists an experiment to test for retroactive inhibition with meaningful discourse would (*a*) use a recall measure instead of, or in addition to, a recognition measure and (*b*) employ test items which would be answered differently on the basis of the first passage than on the basis of the second. The authors know of three studies that fulfill these conditions. Entwisle and Huggins (1964) performed two experiments within the context of an introductory college course in electrical engineering. The 80 students in this course were divided into four mathematical aptitude strata. Within each stratum, half the subjects were randomly assigned to spend 15 minutes studying voltage principles and the next 15 minutes studying the similar current principles. The control group studied voltage principles and then studied irrelevant material on computer programming. The control group made significantly fewer errors than the experimental group on a short-answer test of voltage principles given immediately after interpolated learning. Every item in this test, and the test employed in the second experiment, entailed a competing response from the interpolated learning materials. Following three more lectures, the students were randomly reassigned from within aptitude strata to an experimental and a control group. This time the experimental group studied node-pair and then loop methods. The control group, which studied node-pair solutions and matrix theory, did significantly better than the experimental group on the test of node-pair solutions. The findings from both of these experiments match those obtained with paired-associate lists.

Unfortunately, Hall (1955) has obtained less convincing results. He presented 30 sentences describing a fictitious primitive tribe. One group then received 30 sentences about another tribe, of which 10 sentences involved similar stimuli and conflicting responses. For instance, one of the original learning sentences was "The tribe is located in *southern* Africa." The corresponding interpolated learning sentence was "The tribe is located in *northern* Africa." Following the original learning sentences a second group received irrelevant sentences about the other tribe and a third (control) group completed aptitude tests. The test, which consisted of original learning sentences with blanks in place of underlined words, was administered immediately after original learning, 45 minutes after interpolated learning, and again after 21 days. Considering only the 10 similar-stimulus, competing-response items, none of the groups showed any forgetting 45 minutes after interpolated learning. After 21 days the control group retained 83.4%, the group that had received irrelevant material retained 82.4% and the group that was exposed to competing responses retained 77.5%. While the trend of the data is in the right direction, this is hardly an impressive amount of retroactive inhibition.

In view of the inconsistent findings of the earlier research, the experiments described in this paper were designed to study further the effects of stimulus and response similarity on retroactive inhibition with meaningful discourse. Test questions were divided into three categories. The first category consisted of items for which the stimulus in the original learning passage and the related interpolated learning passage were very similar and the responses were identical. Studying the related interpolated learning passage was expected to facilitate performance on these items. Second were items involving different stimuli and different responses. On items of this type studying the related interpolated learning passage should have no effect or, perhaps, interfere slightly. The third category consisted of items for which reading the related interpolated learning passage should produce strong interference. The stimulus defined by the test item stem, or one very much like it, appeared in both the original learning passage and the related interpolated learning passage; however, the responses associated with this stimulus were conflicting.

The present research also examined the effects of type of test procedure. In the first experiment each test item was presented in both short-answer and multiple-choice form. Investigated, finally, were the effects of giving a test of original learning immediately following the original learning passage. Several studies (cf. Roderick & Anderson, 1968; Spitzer, 1939) indicate that retention at later dates is extraordinarily good when the test covering a meaningful prose passage is administered for the first time as soon as the passage is completed. Presence or absence of an immediate test could influence the amount of retroactive inhibition.

EXPERIMENT I

Method

Passages and tests. Parallel, anthropological discussions of two fictitious, primitive tribes served as the original learning and related interpolated learning passages.[3] These passages covered such topics as climate, topography, religious practices, agriculture, manufacturing skills, and governmental systems. Parallel information was divided into three categories: same, unrelated, and conflicting. For example, both tribes were said to make a kind of beer from squash; for one tribe hunting was discussed, while for the other tribe boat building was covered; and one tribe employed rice as the basic food while wheat was the staple for the other. There were approximately equal amounts of the three types of information. The information varied in kind from very specific (names of food) to general (the family systems). Paralogs of about equal association value (Noble, 1952) were assigned to the two passages in most instances where proper nouns were needed. For example, the name given to one tribe's river was "Brugen"; that assigned to the other tribe was "Kaysen." The unrelated interpolated learning (control) passage was the one on drug addiction used by Ausubel et al. (1968).

The original learning, related interpolated learning, and unrelated interpolated learning passages contained, respectively, 2,240, 2,190, and 1,620 words. A readability analysis using the Farr, Jenkins, and Patterson revision of the Flesch Reading Ease formula (Klare, 1963) gave uncorrected reading levels of 6.9 for the original learning passage, 6.6 for the related interpolated learning passage, and 9.1 for the unrelated interpolated learning passage.

The criterion test on the original learning passage consisted of a 30-item, short-answer test and a matched 30-item multiple-choice test. The items in the two forms of the test contained identical stems, but the multiple-choice test presented four response alternatives whereas the short-answer test contained only blanks or spaces in which to respond. For most multiple-choice items, two foils were chosen from the related interpolated learning passage while one foil and the correct alternative came from the original learning passage. This procedure for selecting foils was employed whenever it resulted in a sensible test item. The short-answer test was graded on the basis of prespecified scoring rules which, in the few items where this was at issue, permitted variations in syntax and unimportant words as long as the "substance" was correct. About 60% of the items could be answered with one or two words. All could be answered in 15 words or less. For the short-answer test subjects recorded their answers in the test booklet; for the multiple-choice test they marked their answers on Digitek blanks. The multiple-choice test was machine scored and corrected for guessing using the formula, $R - W/3$.

There were three types of test items. Facilitation items, so-called because the related interpolated learning passage was expected to facilitate performance on them, contained a stem which was represented in substantially the same form in both the original learning and related interpolated learning passages and which would be answered in the same way on the basis of both passages. Neutral items entailed material which was not discussed in the related interpolated learning passage. The item stem of interference items was represented in both passages; however, the related interpolated learning passage indicated an incorrect, competing response. All of the interference items in the multiple-choice test included among the foils the specific competing response from the related interpolated learning passage. Below is one such item:

> What must a Himootian male do before he is permitted to marry?
>
> A. serve as a Nostow for two weeks
> B. stalk a wild boar
> C. demonstrate progress with his work
> D. obtain a cloak of Buama skin

The original learning passage indicates that before he is permitted to marry the young Himootian male must demonstrate satisfactory progress with his work. The young male of the tribe described in the related interpolated learning passage must complete a feat of daring, such as stalk a wild boar, before he is permitted to marry. One third of the items were of each type.

The Educational Testing Service Wide Range

---

[3] For copies of the two passages and the test, order NAPS Document 01297 from ASIS National Auxiliary Publications Service, c/o CCM Information Sciences, Inc., 909 3rd Avenue, New York 10022; remitting $2.00 for microfiche or $5.00 for photocopies.

Vocabulary Test (French, Ekstrom, & Price, 1963) was used to measure verbal ability.

Design. The experiment involved five factors: presence or absence of the immediate test of original learning; type of interpolated learning passage, related or unrelated; four levels of verbal ability; test response mode, short answer and multiple-choice; and item type, facilitation, neutral, and interference. The latter two variables were within-subject factors. That is, the subjects received both short answer and multiple-choice tests and all three item types. The between-subject factors were completely crossed.

There were two outside control groups. The first of these read the related interpolated learning passage and then immediately completed the original learning test. The second outside control group studied the unrelated interpolated learning passage before taking the original learning test.

Procedure. On Day 1, all subjects in the main part of the experiment read the original learning passage. Immediately thereafter half of them completed the original learning test for the first time. The remaining subjects worked on a short-answer inductive reasoning test and a multiple-choice logical reasoning test, On Day 2, all subjects first took the verbal aptitude test. Then half of those who had and half of those who had not completed the immediate original learning test studied the related interpolated learning passage. The rest received the unrelated interpolated learning material. On Day 8, everyone took the original learning test and filled out a questionnaire inquiring into strategies used to study the materials, the relative care given to each passage, and rehearsal between sessions.

The study was run in six class-sized shifts of 20–30 subjects. All the groups in the experiment were represented in each shift. The subjects were allowed 22 minutes to read each passage, and 12 minutes each for short-answer, multiple-choice, and vocabulary tests; however, a few variations must be noted. On Day 1, the first-period class could be allowed only 7 minutes on the multiple-choice test because "daily announcements" unexpectedly intruded into experimental time. Because of similar difficulties two other classes were afforded only 10 minutes on the multiple-choice test on Day 1. On Days 2 and 8, a few subjects who came late were run apart from the main group and in some cases finished during extra time at the end of the period. These subjects were distributed through the various experimental groups. All or almost all of the subjects had plenty of time to finish all of the tests.

At each stage of the experiment, the experimenter read aloud instructions that were identical for all groups. Where the instructions for the groups were different, subjects were allowed 1 minute to read silently. Directions preceding the passages told the subjects to read carefully, to reread difficult sections, and to be prepared for a short-answer and multiple-choice test. It was stressed that to answer the test questions correctly the subjects would have to understand both general concepts and specific details. Test directions to the outside control groups were augmented to make reasonable the task of taking a test on material not studied.

Subjects. The subjects were 145 students enrolled in six sections of an American History course in a predominantly middle-class, suburban high school. Within each section of the history course, the subjects were randomly assigned to the four main groups and the two outside control groups by the simple expedient of passing out materials stacked in a random order. Students present on Day 1 who were absent on Day 2 or Day 8 were dropped from the study. There seems to have been some differential loss of subjects, perhaps due to the effects of the treatment administered on Day 1. Three subjects who received the immediate original learning test were absent later whereas 13 subjects who did not receive the immediate original learning test failed to show at a subsequent session. Students absent on Day 1 who appeared on Day 2 were assigned to one of the outside control groups.

Results

The mean percentages correct on the parts of delayed posttest for each of the experimental groups appear in Table 1. Since there were disproportionate numbers of cases per cell, an unweighted means analysis of variance was completed. Table 2 contains the analysis of variance summary.

TABLE 1

UNWEIGHTED MEAN PERCENTAGE CORRECT ON THE DELAYED ORIGINAL LEARNING TEST

Group	*N*	Short answer			Multiple-choice		
		Facilitation items	Neutral items	Interference items	Facilitation items	Neutral items	Interference items
Related interpolated learning							
Immediate test	26	63.7	60.4	68.0	68.8	63.7	71.8
No immediate test	21	34.9	39.7	35.9	60.8	58.2	41.8
Unrelated interpolated learning							
Immediate test	27	67.8	70.2	74.6	73.7	84.6	88.8
No immediate test	26	38.3	44.9	48.8	53.4	65.0	77.4

TABLE 2

Analysis of Delayed Original Learning Test Variance, Experiment I

Source	*df*	*MS*	*F*
Between			
Presence of immediate test (T)	1	599.22	56.39**
Interpolated learning passage (IL)	1	129.79	12.21**
Aptitude (A)	3	129.50	12.19**
T × IL	1	.41	.04
T × A	3	26.16	2.46
IL × A	3	10.15	.95
T × IL × A	3	26.75	2.52
Subjects within groups	84	10.62	
Within			
Response mode (R)	1	234.25	105.72**
T × R	1	41.37	18.67**
IL × R	1	11.60	5.23*
A × R	3	14.16	1.88
T × IL × R	3	.64	.28
T × A × R	3	1.06	.48
IL × A × R	3	1.43	.64
T × IL × A × R	3	2.31	1.04
Subjects within groups × R	84	2.26	
Item type (I)	2	14.22	4.95**
T × I	2	5.31	1.84
IL × I	2	30.74	10.70**
T × IL × I	2	15.21	5.29**
A × I	6	2.93	1.02
T × A × I	6	3.52	1.22
IL × A × I	6	5.71	1.99
T × IL × A × I	6	.85	.30
Subjects within groups × I	168	2.87	
R × I	2	.15	.13
T × R × I	2	1.31	1.11
IL × R × I	2	12.57	10.60**
A × R × I	6	.75	.63
T × IL × R × I	2	4.77	4.02*
T × A × R × I	6	2.88	2.42*
IL × A × R × I	6	2.68	2.26*
T × IL × A × R × I	6	2.03	1.71
Subjects within groups × R × I	168	1.19	

* $p < .05$.
** $p < .01$.

Comparison of control groups. The purpose of comparing the control groups was to determine whether the related passage affected performance on the facilitation, neutral, and interference items in the intended manner. Figure 1 shows that it did. There was a significant Passage × Item Type interaction, $F = 72.76$, $df = 2/86$, $p < .01$. The mean for the related control on interfering items is below zero because multiple-choice items were corrected for "guessing."

Effects of interpolated learning passage and item type. Turning now to the analysis of the retention of the experimental groups, the subjects who received the unrelated interpolated learning passage scored significantly higher on the delayed test of original learning than those who received the related interpolated learning passage, $F = 12.21$, $df = 1/84$, $p < .01$. A significant Interpolated Learning Passage × Item Type interaction was also obtained, $F = 10.70$, $df = 2/168$, $p < .01$. As Figure 2 shows, the interaction takes roughly the form that would be expected on the basis of rote verbal learning research.

The group that received the related interpolated learning passage should have done much better on the facilitating items than on the interfering items. The fact that it did not is attributable to differences in level of learning and/or the difficulty of the test items. On the immediate test, the average percentages were 64.2, 70.9, and 78.6 on the facilitating, neutral, and interfering items, respectively.

The failure of the related interpolated learning group to surpass the unrelated interpolated learning group on the facilitating items was also unexpected. However, the level of learning was somewhat higher in the latter group. In order to adjust for differences in level of learning, the percentage retained, defined as 100 times the ratio of the mean delayed test score to the mean immediate test score, was computed. For the two groups that did not get the immediate test, the average of the means of the remaining two groups was used as the estimate of learning. Figure 3 contains percentage retained as a function of interpolated learning passage and item type. The data

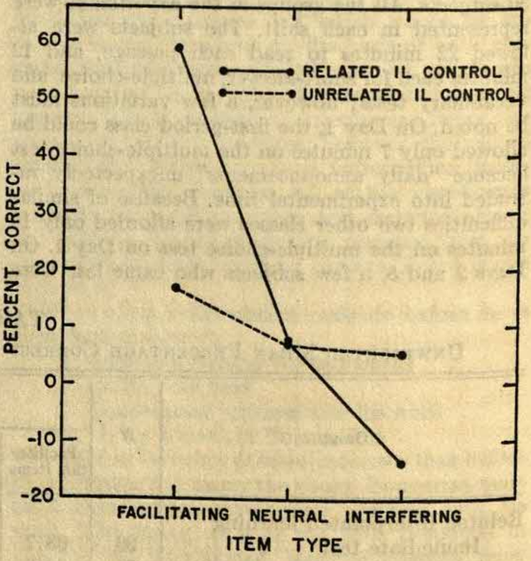


Fig. 1. Percentage correct on the original learning test by item type for the outside control groups.

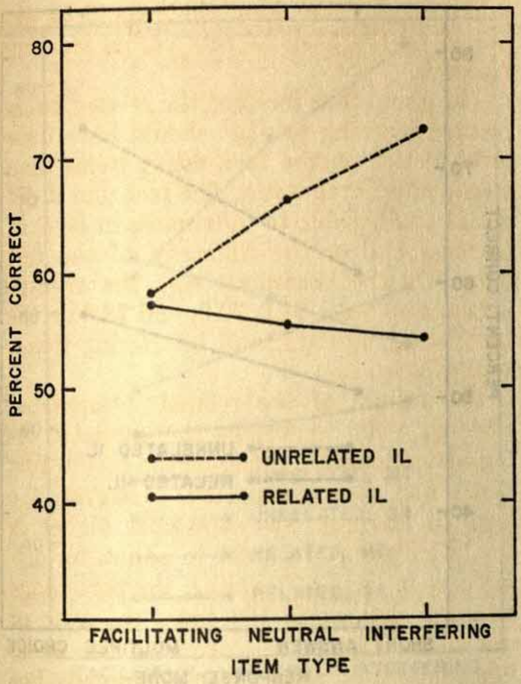


Fig. 2. Percentage correct on the delayed original learning test as a function of interpolated learning passage and item type.

now fit closely expectations based on the paired-associate literature.

Response mode and presence of immediate test. Receiving the immediate test of original learning produced a substantial increment in delayed test performance, $F = 56.39$, $df = 1/84$, $p < .01$. The unweighted delayed test mean for groups that received the immediate test was 71.4% while the comparable figure for groups that did not receive the immediate test was 49.9%.

Performance was much higher on the multiple-choice than short-answer portion of the delayed test, $F = 105.72$, $df = 1/84$, $p < .01$, despite the fact that the multiple-choice items were corrected for guessing. Figure 4 represents the Presence of Immediate Test × Response Mode interaction, $F = 18.67$, $df = 1/84$, $p < .01$. This figure shows that the immediate test facilitated performance more on short-answer items then on multiple-choice items.

Both the immediate test and response mode interacted with interpolated learning passage and item type to affect the amount of retroactive inhibition. The significant, F

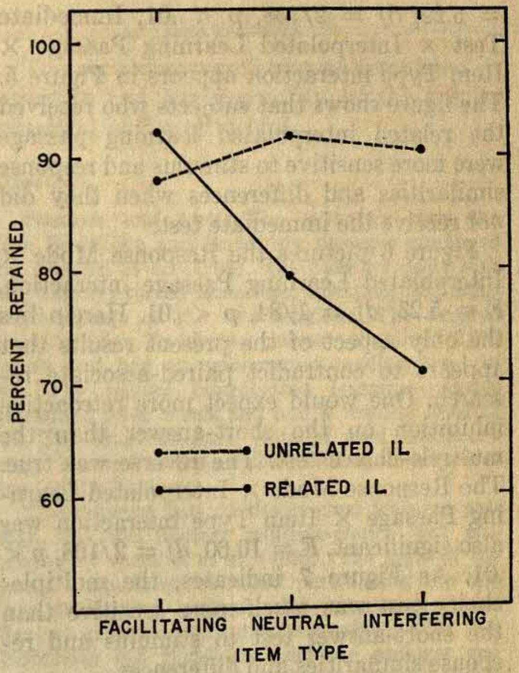


Fig. 3. Percentage retained as a function of interpolated learning passage and item type.

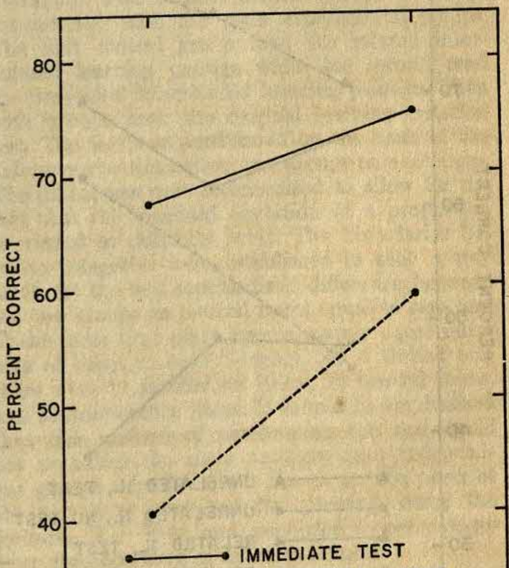


Fig. 4. Percentage correct on the delayed original learning test as a function of presence or absence of the immediate test and response mode.

$= 5.29$, $df = 2/168$, $p < .01$, Immediate Test × Interpolated Learning Passage × Item Type interaction appears in Figure 5. The figure shows that subjects who received the related interpolated learning passage were more sensitive to stimulus and response similarities and differences when they did *not* receive the immediate test.

Figure 6 pictures the Response Mode × Interpolated Learning Passage interaction, $F = 5.23$, $df = 1/84$, $p < .01$. Herein lies the only aspect of the present results that appears to contradict paired-associate research. One would expect more retroactive inhibition on the short-answer than the multiple-choice test. The reverse was true. The Response Mode × Interpolated Learning Passage × Item Type interaction was also significant, $F = 10.60$, $df = 2/168$, $p < .01$. As Figure 7 indicates, the multiple-choice test was much more sensitive than the short-answer test to stimulus and response similarities and differences.

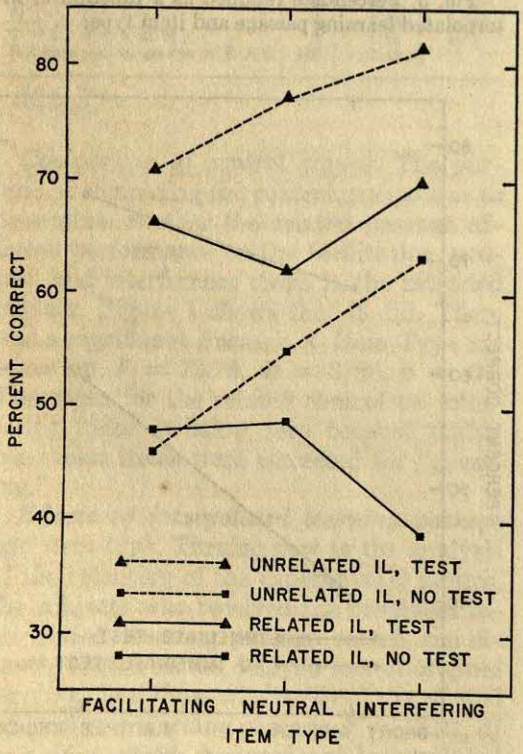


FIG. 5. Percentage correct on the delayed original learning test as a function of presence or absence of the immediate test, interpolated learning passage, and item type.

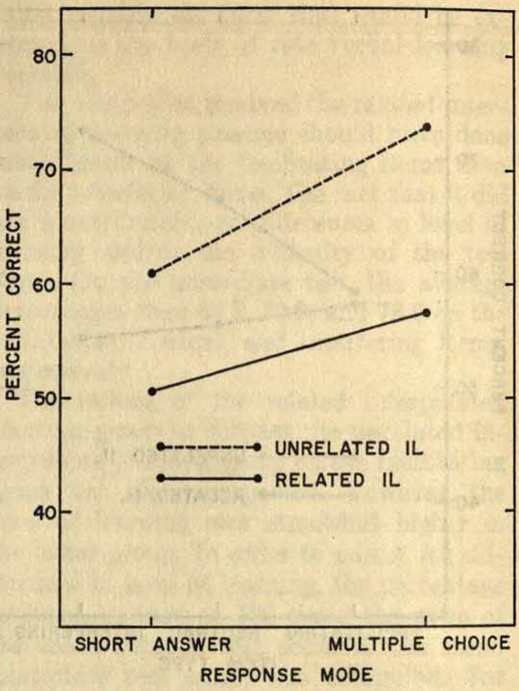


FIG. 6. Percentage correct on the delayed original learning test as a function of response mode and interpolated learning passage.

Subject strategies. There were several significant ($\alpha = .05$) correlations between answers to items on the postexperiment questionnaire and performance on the delayed test. Summarized here are the means of the correlations of the various groups. Performance correlated .24 with reported carefulness in studying the first passage and .28 with reported carefulness in studying the second passage. It might have been expected that carefulness in studying the related passage would be positively correlated with performance on facilitation items and negatively correlated with performance on interference items; however, both correlations were positive and of the same magnitude. In fact, correlations between performance and reported care on the unrelated passage were comparable.

When asked which passage they studied most carefully, 60% indicated the original learning passage, 32% said they studied the passages with equal care, while only 8% said they studied the interpolated learning passage more carefully than the original

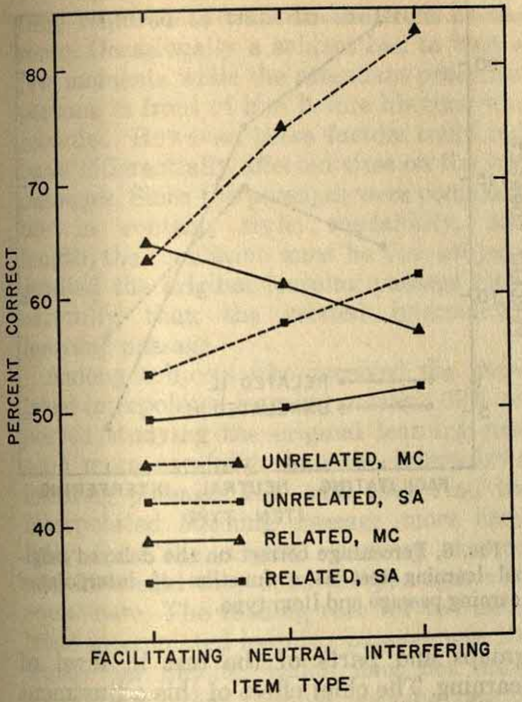


Fig. 7. Percentage correct on the delayed original learning test as a function of response mode, interpolated learning passage, and item type.

learning passage. The percentages were the same whether the subject received the related or the unrelated interpolated learning passage. The extent to which the subjects reported discussing the ideas in the original learning passage, or the questions in the immediate original learning test with friends, correlated .23 with delayed test scores.

One question asked the extent to which the subject tried to remember what the first booklet said, while studying the second, in order to get clear on similarities and differences between the two. Among those who received the related passage, answers to this question correlated .28 with delayed test scores. Finally, delayed test performance correlated .18 with reported willingness to participate again in another, similar experiment.

Other effects. Verbal ability was significantly related to recall, $F = 12.19$, $df = 3/84$, $p < .01$. There were also three four-way interactions, two of which involved level of verbal ability, significant at the .05 level. These were left uninterpreted.

Experiment II

The second experiment was completed in order to test the generality of the results obtained in Experiment I.

Method

Passages and tests. The materials were the Zen Buddhism and Buddhism passages employed by Ausubel et al. (1968). The passages dealt with the history, sacred literature, doctrine, and ethical teachings of the two religions. The original learning (Zen Buddhism) passage contained 2,990 words and had an uncorrected reading level of 6.7. The related interpolated learning passage contained 2,640 words and had a reading level of 6.9. The drug addiction material used in Experiment I served as the unrelated interpolated learning (control) passage.

The criterion test on the original learning passage was the 35-item multiple-choice test employed by Ausubel et al. (1968). The logic of the present experiment required partitioning the test into facilitating, neutral, and interfering items. Several graduate students unacquainted with Zen Buddhism and Buddhism read the passages and then categorized the test items. Unfortunately there was only a moderate degree of agreement among the raters. The decision was made, therefore, to partition the test using an a posteriori technique. Two outside control groups were run concurrently with the main experimental groups. The first control group read the related interpolated learning passage while the second read the unrelated interpolated learning passage. Then both groups took the original learning criterion test. The test was partitioned on the basis of the differences between these two groups on each item. The differences were standardized to allow for the fact that the standard deviation of a proportion is related to difficulty level. The boundaries between categories were established in such a way as to set the net standardized difference between the two groups on neutral items equal to zero and at the same time place approximately equal numbers of items in each category. As it turned out, there were 10 facilitation items, 12 neutral items, and 13 interference items. It should be emphasized that this method of partitioning the test could not prejudice the main analysis since independent groups were involved. Scores on the parts of the test were corrected for guessing, using the formula, $R - W/4$, and converted to percentages since the parts were of different lengths.

Design. The factors in the experiment were presence or absence of the immediate test of original learning, type of interpolated learning, passage—related or unrelated, and item type—facilitating, neutral, and interfering. In all other respects the design was the same as the design of Experiment I.

Subjects and procedures. The subjects were 152 juniors and seniors from a high school in a middle-class area in a Midwestern industrial city of 90,000. Eighteen additional subjects present on Day 1 were dropped from the study because of absences on Day 2 or Day 8.

The study was completed in the school auditorium in three shifts of 50–60 students. On Day 1 subjects were assigned to read the original learning passage or one of the two control passages by passing out materials stacked in a random order. The study was subject-paced. When a subject finished studying a passage, he reported to an attendant at the front of the room who recorded the subject's name and the time. Of each block of four subjects finishing the original learning passage, one subject was randomly assigned to each of the four experimental groups. This procedure equated the original learning study time of the four groups. In all other respects the procedure was the same as in Experiment I.

Results

The unweighted means analysis of delayed original learning test variance summarized in Table 3 revealed three significant effects. Theoretically most important was the Interpolated Learning Passage × Item Type interaction, $F = 3.91$, $df = 2/174$, $p < .05$, which is graphed in Figure 8. While the results are not as strong or as orderly as those obtained in Experiment I, the figure does show that, as expected, the related interpolated learning group did better than the unrelated interpolated learning group on facilitation items but worse on interference items. Unlike Experiment I, the related interpolated learning group also did better on neutral items. Figure 9 contains another version of the Interpolated Learning Passage × Item Type interaction, with percent retained (computed in the manner described earlier) as the dependent variable in order to adjust for variations between groups and parts of the test in level of learning. The chief effect of this adjustment was to exaggerate the difference between conditions on the neutral items.

TABLE 3

ANALYSIS OF DELAYED ORIGINAL LEARNING TEST VARIANCE, EXPERIMENT II

Source	*df*	*MS*	*F*
Between			
Immediate test (T)	1	.07	.95
IL passage (IL)	1	.05	.67
T × IL	1	.18	2.32
Subjects within groups	87	.08	
Within			
Item type (I)	2	.29	11.58**
T × I	2	.11	4.18*
IL × I	2	.10	3.91*
T × IL × I	2	.01	.25
Subjects within groups × I	174	.03	

* $p < .05$.
** $p < .01$.

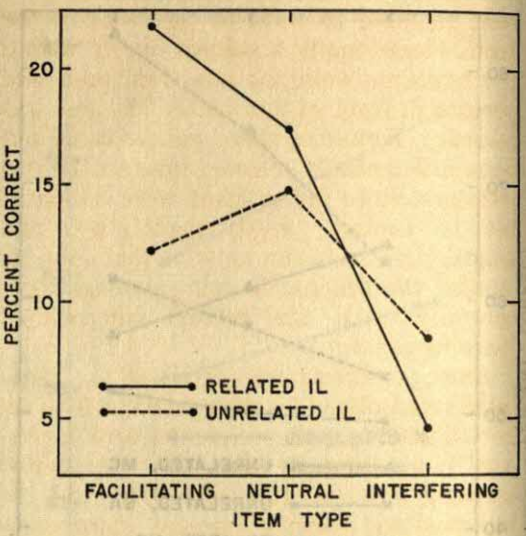


FIG. 8. Percentage correct on the delayed original learning test as a function of interpolated learning passage and item type.

Item type had a significant effect, $F = 11.58$, $df = 2/174$, $p < .01$. The unweighted means were 16.9%, 16.0% and 6.5%, respectively, for the facilitation, neutral, and interference items. The Immediate Original Learning Test × Item Type interaction, $F = 4.18$, $df = 2/174$, $p < .05$, is pictured in Figure 10. We have no explanation for this interaction.

When subjects who had received the related interpolated learning passage were asked on the postexperiment questionnaire which passage they had studied most carefully, 59% indicated the original learning passage, 30% said they studied the two passages with equal care, while only 11% said they studied the interpolated learning passage more carefully than the original learning passage. An analysis of the time data confirmed the subjects' self-reports. The reading rate on the original learning passage was 130 words per minute whereas the rate on the related interpolated learning passage was 170 words per minute, $t = 7.29$, $df = 37$, $p < .01$. It should be noted that the estimates of reading rates are not very accurate. Included in reading time was the time required to walk to the front of the room. Occasionally a subject had to wait a few moments while the attendant processed persons in front of him before his time was recorded. However, these factors could not have differentially affected time on the two passages. Since the passages were comparable in content, style, readability, and length, the conclusion must be that subjects studied the original learning passage more carefully than the related interpolated learning passage.

Among subjects who received the unrelated interpolated learning passage, 39% reported studying the original learning passage more carefully than the interpolated learning passage, 41% said they read the interpolated learning passage more carefully than the original learning passage, and 20% said they read the passages with equal care. The reading rate for the unrelated interpolated learning passage was 132 words per minute, which does not differ from the average original learning rate of 130 words per minute.

There were small positive correlations ($r \leqslant .15$) between performance on the delayed original learning test and reported care in studying the original learning passage, reported care in studying the interpolated learning passage, and the extent to which subjects reported discussing the original learning passage and/or immediate original learning test with his classmates. For groups that got the related interpolated learning passage, delayed test performance correlated .45 with the extent to which subjects reported thinking about similarities and differences between the passages while studying the interpolated learning passage.

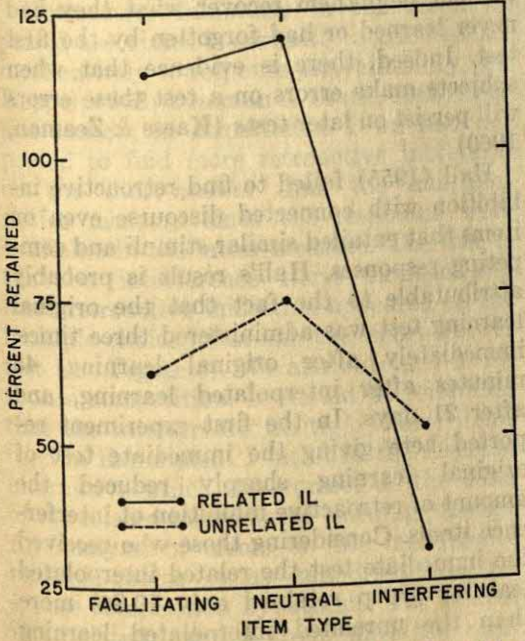


Fig. 9. Percentage retained as a function of interpolated passage and item type.

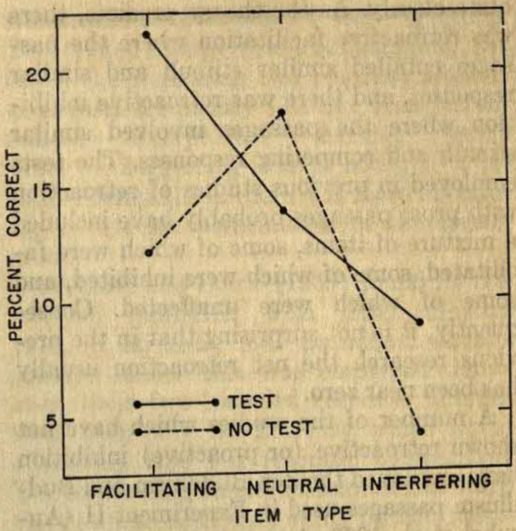


Fig. 10. Percentage correct on the delayed original learning test as a function of presence or absence of the immediate test and item type.

Ratings of how interesting the original learning passage was showed an average correlation of .34 with delayed original learning test performance. The correlations between ratings of the interpolated learning passage and performance on the delayed test averaged .32 for the groups that received the related interpolated learning passage and .08 for the groups that got the unrelated interpolated learning passage. There was an average correlation of .22 between delayed test performance and reported willingness to participate again in another, similar experiment.

Discussion

The data from both of the experiments reported in this paper fit interference theory quite closely. As the theory predicts, there was retroactive facilitation where the passages entailed similar stimuli and similar responses, and there was retroactive inhibition where the passages involved similar stimuli and competing responses. The tests employed in previous studies of retroaction with prose passages probably have included a mixture of items, some of which were facilitated, some of which were inhibited, and some of which were unaffected. Consequently, it is not surprising that in the previous research the net retroaction usually has been near zero.

A number of the studies which have not shown retroactive (or proactive) inhibition have employed the Zen Buddhism and Buddhism passages used in Experiment II (Ausubel et al., 1968; Ausubel, Stager, & Gaite, 1969; Gaite, 1968). In addition to the mixing of different types of test items, a second difficulty which probably weighed against finding inhibition is the very low level of learning obtained with the Zen Buddhism passage and accompanying test. In the second experiment reported here the unweighted mean correct on the immediate Zen Buddhism test was 17.2% (corrected for guessing).[4]

The mean recall of the group which read the unrelated (control) interpolated learning passage was so poor on interference items, only 8.4%, that there was not room for much of the predicted retroactive inhibition in the related interpolated learning condition to show itself. On the other hand there was plenty of room on the upside for facilitation to appear. We believe that the low level of learning was the chief reason why stronger effects failed to appear in Experiment II and a major reason why Ausubel and his associates have consistently failed to find proactive or retroactive inhibition using the Zen Buddhism materials.

One would expect either no effect or a small amount of retroactive inhibition on neutral items. As a matter of fact Experiment I showed retroactive inhibition on neutral items whereas retroactive facilitation appeared in Experiment II. Experiment I is probably the more trustworthy of the two experiments. In addition to the difficulty discussed in the preceding paragraph, a relatively crude procedure was employed to identify item types in Experiment II.

In Experiment I, taking the immediate test of original learning substantially improved later recall. This is not an isolated finding. It is well established that responding to questions during (Michael & Maccoby, 1961; Rothkopf, 1966) or immediately after (Roderick & Anderson, 1968; Spitzer, 1939) studying an instructional communication strongly facilitates later recall even when, as in the present case, no knowledge of results is provided. The question is why in Experiment II receiving the immediate test of original learning failed to improve performance on the delayed test. We attribute the failure to find facilitation to the low level of learning. Spitzer (1939) found that giving a test at various intervals after subjects read a passage helped them to remember on later tests what they could recall up to the point of the first test but did not help them recover what they had never learned or had forgotten by the first test. Indeed, there is evidence that when subjects make errors on a test these errors will persist on later tests (Kaess & Zeaman, 1960).

Hall (1955) failed to find retroactive inhibition with connected discourse even on items that entailed similar stimuli and competing responses. Hall's result is probably attributable to the fact that the original learning test was administered three times, immediately after original learning, 45 minutes after interpolated learning, and after 21 days. In the first experiment reported here giving the immediate test of original learning sharply reduced the amount of retroactive inhibition of interference items. Considering those who received the immediate test the related interpolated learning group retained only 10.8% more than the unrelated interpolated learning group, $t = 1.84$, $df =$ (approx.) 42, $p > .05$, whereas the difference between the two interpolated learning conditions was 30.8% for subjects who did not receive the immediate original learning test, $t = 4.88$, $df =$ (approx.) 42, $p < .01$.

---

[4] Learning has been somewhat higher in the previous experiments using the Zen Buddhism passage. However, in these studies scores were not corrected for guessing and a shortened version of the Zen Buddhism test was used from which four difficult items were removed.

There was an interaction in Experiment I between taking the immediate test and the response mode of the delayed test. Performance on short-answer items was more strongly facilitated than performance on multiple-choice items. This is a finding in which we have considerable confidence, since Roderick and Anderson (1968) observed the same phenomenon in a study involving recall of information about classical conditioning. The usual assumption is that a subject must have acquired and remember an associative connection to answer either a short-answer or multiple-choice item. The big difference is that to answer a short-answer item the response must be available in the subject's repertoire, whereas the response is provided for him in a multiple-choice question. The fact that performance on short-answer items is more strongly facilitated than performance on multiple-choice items seems to indicate, therefore, that completing an immediate test primarily influences response availability rather than specific associative connections.

Retroactive inhibition in the recall of paired associates is currently attributed almost entirely to response unavailability (cf. Keppel, 1968; Postman & Stark, 1969). For this reason, in Experiment I we had expected to find more retroactive inhibition on the short-answer than the multiple-choice measure since the responses were available in the latter measure. The fact is that on interference items substantially more retroactive inhibition appeared on the multiple-choice than on the short-answer test (see Figure 6). We attribute this to response competition. It should be reiterated that the interference multiple-choice items used in Experiment I included the response alternative which was correct on the basis of the related interpolated learning passage. An analysis of errors on the interference multiple-choice items supported the response competition hypothesis. The related interpolated learning group chose the competing interpolated learning response 17.7% of the time while the figure was 4.3% for the unrelated interpolated learning group. The related interpolated learning group also chose the remaining two response alternatives more frequently (17.0%) than the unrelated interpolated learning group (10.4%), suggesting that in addition to response competition there was more unlearning of specific associations in the former group.

Postman and Stark (1969) found no retroactive inhibition with letter-adjective pairs in an A-B, A-C paradigm when a multiple-choice test was employed. However, their test included response alternatives from only the first list, which would effectively rule out response competition. The contrasting results of Experiment I and the Postman and Stark study do not, therefore, indicate that different processes are involved in retroactive inhibition with prose and retroactive inhibition with paired associates. It can be argued that Postman and Stark overestimated the importance of response availability because they failed to consider the role of response competition.

While the present studies have illustrated that it is possible to obtain consistent, interpretable results outside of the confines of the laboratory, several difficulties with free-field research on memory became apparent. Most serious was the strong tendency, revealed on the postexperiment questionnaires after both experiments and in the reading rate data from Experiment II, of subjects to study the interpolated learning passage less carefully than the original learning passage. Naturally, this could attenuate the effects of interpolated learning. We are unable to think of a satisfactory way to control the level of learning of lengthy prose passages. One possibility would be to give tests after each passage as a basis for statistical control. However, giving tests will itself have strong effects, as Experiment I demonstrated. Merely equating total study time is inadequate since people cannot be forced to make good use of available time.

A less serious problem was the reported tendency of some subjects to discuss the original learning passage and immediate original learning test with their friends. Those who did improved their recall, as evidenced by the small positive correlations between answers to the question and performance on the delayed original learning test.

Subjects in both experiments who reported thinking about the similarities and differences between the original learning and related interpolated learning passages while studying the interpolated learning passage performed somewhat better on the delayed original learning test. These data suggest that retroactive inhibition might be reduced and potential facilitative effects enhanced if the materials expressly contrasted the similarities and differences of consecutive passages (cf. Ausubel, 1963).

REFERENCES

AUSUBEL, D. P. *The psychology of meaningful verbal learning*. New York: Grune & Stratton, 1963.

AUSUBEL, D. P., ROBBINS, L. C., & BLAKE, E. Retroactive inhibition and facilitation in the learning of school materials. *Journal of Educational Psychology,* 1957, **48,** 334–343.

AUSUBEL, D. P., STAGER, M., & GAITE, A. J. H. Retroactive facilitation in meaningful verbal learning. *Journal of Educational Psychology*, 1968, **59,** 250–255.

AUSUBEL, D. P., STAGER, M., & GAITE, A. J. H. Proactive effects in meaningful verbal learning and retention. *Journal of Educational Psychology,* 1969, **60,** 59–64.

ENTWISLE, D. R., & HUGGINS, W. H. Interference in meaningful learning. *Journal of Educational Psychology,* 1964, **55,** 75–78.

FRENCH, J. W., EKSTROM, R. B., & PRICE, L. A. *Kit of Reference Tests for Cognitive Factors.* Princeton, N. J.: Educational Testing Service, 1963.

GAITE, A. J. H. A study of retroactive inhibition and facilitation in meaningful verbal learning. Unpublished doctoral dissertation, University of Toronto, 1968.

HALL, J. F. Retroactive inhibition in meaningful material. *Journal of Educational Psychology,* 1955, **46,** 47–52.

KAESS, W., & ZEAMAN, D. Positive and negative knowledge of results on a Pressey-type punchboard. *Journal of Experimental Psychology,* 1960, **60,** 12–17.

KEPPEL, G. Retroactive and proactive inhibition. In T. R. Dixon and D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

KING, D. J., & COFER, C. N. Retroactive interference in meaningful material as a function of the degree of contextual constraint in the original and interpolated learning. *Journal of General Psychology,* 1960, **63,** 145–158.

KLARE, G. C. *The measurement of readability.* Ames: Iowa University Press, 1963.

MCGEOCH, J. A., & MCKINNEY, F. The susceptibility of prose to retroactive inhibition. *American Journal of Psychology,* 1934, **46,** 429–436.

MCGOVERN, J. B. Extinction of associations in four transfer paradigms. *Psychological Monographs,* 1964, **78** (16, Whole No. 593).

MEHLER, J., & MILLER, G. A. Retroactive interference in the recall of simple sentences. *British Journal of Psychology,* 1964, **55,** 295–301.

MICHAEL, D. N., & MACCOBY, N. Factors influencing the effects of student participation on verbal learning from films: Motivating versus practice effects, "feedback," and overt versus covert responding. In A. A. Lumsdaine (Ed.), *Student response in programmed instruction.* Washington, D. C.: National Academy of Science, National Research Council, 1961.

NOBLE, C. E. An analysis of meaning. *Psychological Review,* 1952, **59,** 421–430.

OSGOOD, C. E. The similarity paradox in human learning. *Psychological Review,* 1949, **56,** 132–143.

POSTMAN, L., & STARK, K. Role of response availability in transfer and interference. *Journal of Experimental Psychology,* 1969, **79,** 168–177.

RODERICK, M. C., & ANDERSON, R. C. Programmed introduction to psychology versus textbook-style summary of the same lesson. *Journal of Educational Psychology,* 1968, **59,** 381–387.

ROTHKOPF, E. Z. Learning from written instructive material: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal,* 1966, **3,** 241–249.

SLAMECKA, N. J. Retroactive inhibition of connected discourse as a function of practice level. *Journal of Experimental Psychology,* 1960, **59,** 104–108.

SLAMECKA, N. J. Retention of connected discourse as a function of duration of interpolated learning. *Journal of Experimental Psychology,* 1962, **63,** 480–486.

SPITZER, H. F. Studies in retention. *Journal of Educational Psychology,* 1939, **30,** 641–656.

(Received January 11, 1970)

You don't have to use a text with this reader

Before Glock, you had no choice, A text was a *must*.

Now there's Glock's READINGS IN EDUCATIONAL PSYCHOLOGY.

The variety and sparkle of a reader. But the thematic unity and continuity of a text.

So you don't have the job of selecting the right supplementary material, then fleshing out the background your students need to understand the articles.

Glock's done it all for you.

He's organized 18 chapters under one theme: "the teacher as a hypothesis-maker." Relating every issue in the field to the total teaching/learning process. Supplying lucid points of view, on every topic.

And he does it all with the sharpest writings of the last ten years. Including choice innovations—excerpts from popular non-texts by John Holt, Belle Kaufman, Herbert Kohl, and others.

With introductions so intensive and thorough, students can link all the ideas without a hitch.

Glock—the reader that stands alone.

READINGS IN EDUCATIONAL PSYCHOLOGY
Edited by Marvin D. Glock, Cornell University
1971 Approx. 480 pages In press

For more information, contact your local Wiley representative or write Wayne Anderson, Promotion Manager.

wiley

JOHN WILEY & SONS, Inc.
605 Third Avenue, New York, N.Y. 10016
In Canada: 22 Worcester Road, Rexdale, Ontario.

FROM HARPER & ROW

READINGS IN LEARNING AND HUMAN ABILITIES: Educational Psychology, Second Edition

Edited by RICHARD E. RIPPLE

An excellent supplement to any introductory educational psychology text, this collection of 41 unabridged readings—most of them dealing with empirical research—introduces the student to original literature in the field. Three-quarters of the material postdates publication of the first edition. Sectional introductions /interpretive headnotes for each reading.

484 pp., $6.50. Paper. March, 1971.

EDUCATION AND MENTAL HEALTH

LOUIS KAPLAN, Assistant Superintendent, Torrance (Calif.) City Schools

Adapted from the author's earlier work, MENTAL HEALTH AND HUMAN RELATIONS IN EDUCATION, this new text assembles and interprets materials from psychology, education, sociology, psychiatry, pediatrics, anthropology, and related behavioral sciences. Based on the theory that schools have an important function beyond inculcation of skills and knowledge, the text surveys research, case materials, and practical experience to provide educators with the information they need to create a more humanistic school environment.

522 pp.; $9.95. January, 1971.

BASIC STATISTICAL METHODS, Third Edition

N. M. DOWNIE, Purdue University, and ROBERT W. HEATH, Stanford University

This introductory text stresses computation, interpretation, and application of frequently encountered statistics. Designed for nonmathematics majors, the orientation is practical rather than theoretical. Features of the Third Edition: change of emphasis from hand operations to data control by calculator or by computer; inclusion of a double classification analysis in analysis of variance techniques; revised and expanded treatment of correlations. Study Guidebook ($2.95).

356 pp.; $9.95. 1970.

ANNOUNCING THE THIRD EDITION OF . . .

LEARNING AND HUMAN ABILITIES: Educational Psychology

HERBERT J. KLAUSMEIER, The University of Wisconsin, and
RICHARD E. RIPPLE, Cornell University

This leading text presents a systematic unification of the field from the perspective of its primary concerns: the nature, conditions, outcomes, and evaluation of classroom learning. With a realistic awareness of how educational principles can best be *applied today*, the Third Edition takes into account the negative effects on education of war, drugs, violence, racism, and poverty. Rewritten for clarity and readability, the text incorporates the latest empirical research studies. Among the organizational changes: individual differences, instructional materials, and motivation are covered earlier in the book; factual information and concepts, formerly one chapter, are now treated in two expanded chapters. Internal study aids include an advance organizer (outline of key concepts) for each chapter, review-discussion questions at the end of each main section of each chapter, "flagging" (by special brackets) of important within-text passages, and annotated chapter-end suggestions for further reading. Profusely illustrated with charts and diagrams. *File of Test Items* (multiple-choice questions) for the instructor by Ruth S. Nickse, Cornell University, and Richard E. Ripple.

810 pp.; $11.95. March, 1971.

STUDENT GUIDE

by Ruth S. Nickse and Richard E. Ripple

Contains, for each chapter in the text, an outline of the chapter, a multiple-choice self-test, discussion questions, and projects; also, answers to the self-tests, supplementary reading list of general books, and annotated list of professional journals.

148 pp.; $3.95. Paper. March, 1971.

1817

HARPER & ROW, PUBLISHERS

49 EAST 33D STREET, NEW YORK 10016

FROM HARPER & ROW . . . NEW PERSPECTIVES ON DEVELOPMENT AND LEARNING

EDUCATIONAL RESEARCH BY PRACTITIONERS: An Elementary Casebook

WILLIAM C. BUDD and SAM P. KELLY, Western Washington State College

The only book in the field which examines actual studies and investigations typically undertaken by teachers and other school personnel. Cases are accompanied by critical commentaries on their necessity, design, and procedures. There are also unevaluated cases which may be used as exercises. *148 pp.; $2.95. Paper. 1970.*

HUMAN DEVELOPMENT: From Birth Through Adolescence, Second Edition

IRA J. GORDON, University of Florida

Scientific and theoretical information about psychological, social, and physical development from birth to age 20 is here integrated into a consistent theory of self-development. Instructor's Manual. *408 pp.; $8.95. 1969.*

INTRODUCTION TO HUMAN MEMORY: A Psychological Approach

MICHAEL J. A. HOWE, University of Alberta

A concise introduction to the knowledge that experimental psychologists have gained about human memory during the past few years, considering basic questions concerning the nature and utility of the concept of memory, relating it to other functions, and demonstrating how problems of memory research have been approached. *113 pp.; $2.95. Paper. 1970.*

WRITING BEHAVIORAL OBJECTIVES: A New Approach

H. H. McASHAN, University of Florida

An effective guide for development of behavioral and other types of performance objectives through a goals-setting approach. Promotes objective-writing competency in both instructional and noninstructional areas of education. The author incorporates a variety of learning strategies associated with programmed learning. *116 pp.; $2.50. Paper. 1970.*

STUDY GUIDE to accompany Mussen, Conger, and Kagan's CHILD DEVELOPMENT AND PERSONALITY, Third Edition

FAY-TYLER M. NORTON, Cuyahoga Community College

Organized to correspond chapter by chapter with the leading text in the field, CHILD DEVELOPMENT AND PERSONALITY, this study guide offers directions for approaching the reading of each chapter; an overview of each chapter in the form of questions; content review with content review questions; various types of short-answer questions, with answers; and problems relating the text to life and to the professions concerned with child growth and development. *Tentative: 214 pp.; $2.95. Paper. February, 1971.*

CHILD DEVELOPMENT AND PERSONALITY, Third Edition

by Mussen, Conger, and Kagan
795 pp.; $10.95. 1969.

READINGS IN CHILD DEVELOPMENT AND PERSONALITY, Second Edition

edited by Mussen, Conger, and Kagan
595 pp.; $5.95. Paper. 1970.

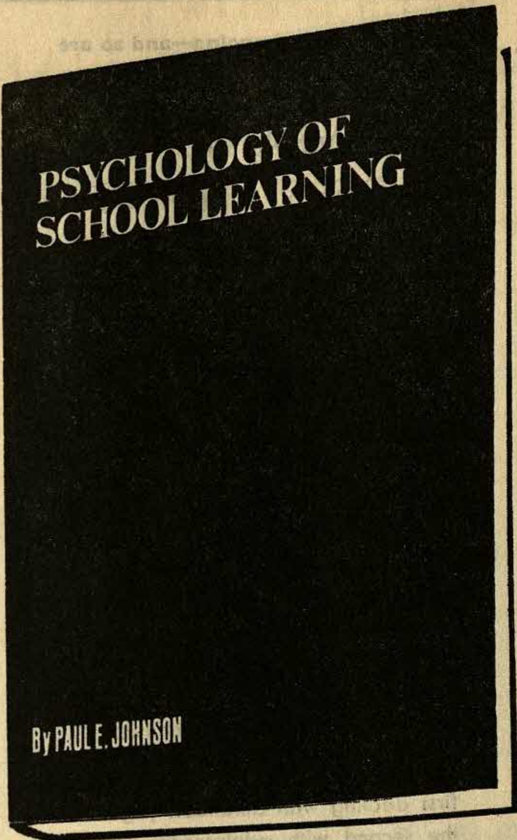
CHILD LEARNING, INTELLIGENCE, AND PERSONALITY: Principles of a Behavioral Interaction Approach

ARTHUR W. STAATS, University of Hawaii

Outlining an approach in which traditional conflicts between personality theory and behavioral theory are resolved, this book presents a new and challenging theory of the development of intelligence and other aspects of personality and their function in subsequent adjustment and behavior modification. Throughout, the author repeatedly relates both his general conception of child development and his specific analyses to the measurement of intelligence, thus achieving an unusual rapprochement among principles of learning, the conception of intelligence, and the measurement of intelligence. *358 pp.; $8.95. January, 1971.*

Send for our 1971 catalogs: education/ #CT 36; psychology/ #CT 43

Harper & Row, Publishers
1817
49 E. 33d St., N.Y. 10016



Puts learning theory to work

What's the use of teaching learning theory if your students can't put it in action when they start to teach?

With JOHNSON's text they can.

It shows them how to effect learning and behavioral change *in the classroom.*

Not with prescriptions or illustrative anecdotes. But with instances of general behavioral models to back up what your students know intuitively about behavior.

With transfer designs to determine what students learn.

And an annotated list of suggested readings.

What students learn from PSYCHOLOGY OF SCHOOL LEARNING by JOHNSON, they put into practice.

PSYCHOLOGY OF SCHOOL LEARNING
By PAUL E. JOHNSON, University of Minnesota
1971 Approx. 144 pages $5.50

For further information, contact your Wiley representative or write Wayne Anderson, Promotion Manager.

wiley
JOHN WILEY & SONS, Inc., 605 Third Avenue, New York, N.Y. 10016
In Canada: 22 Worcester Road, Rexdale, Ontario

the times are changing—and so are children

A SYMPATHETIC UNDERSTANDING OF THE CHILD: Six to Sixteen

by DAVID ELKIND
University of Rochester

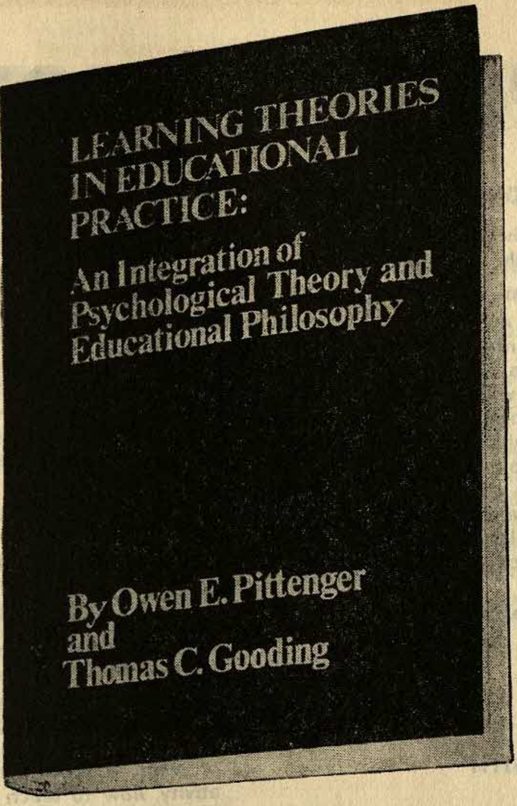
The author, a Piagetian with many years' experience as teacher and therapist, writes with insight and understanding about the major aspects of child and adolescent development today. He focuses on children in a social context, particularly children in the schools and the schools' influence on their behavior.

The book is divided into two sections, the first dealing with childhood (ages 6–11), the second with adolescence (ages 12–16). For each age group, he gives brief, informal descriptions of the continuum of development; facets of growth and development; social-cultural change; general characteristics; the world of self, home, community; cognitive development; the age group in relation to adults; and adults in relation to the age group. Examples from real life add interest and humor to these discussions.

Ideally suited for courses in Child Psychology and Educational Psychology. 1971, 5 7/16 x 8 3/8, 160 pp., paperbound

now available from

ALLYN & BACON, INC.
College Div., Dept. 893
470 Atlantic Ave., Boston, MA
02210



A Meeting of Minds

Psychological theory meets educational philosophy in Pittenger/Gooding.

Three formulations of educational philosophy (traditional, technological, and transactional) meet four learning theories (Thorndike, Skinner, Gestalt, and Combs and Snygg).

And the minutes are in the form of a clear, authoritative exposition.

For students who wish to develop their own critical appraisal of the meeting, an "Instant Comparison" guide in the form of summary tables is also present.

But developmental psychology is absent. Because its ideas are met and explained in so many other current texts.

Pittenger/Gooding—meeting the need for reestablishing education as a personally and socially significant experience.

Learning Theories in Educational Practice: An Integration of Psychological Theory and Educational Philosophy. By Owen E. Pittenger and Thomas C. Gooding, both at the State University of New York at Oswego. 1971. 256 pages. $7.50. For more information, contact your local Wiley representative or write Wayne Anderson, Promotion Manager.

wiley **John Wiley & Sons, Inc.**
605 Third Avenue, New York, N.Y. 10016
In Canada: 22 Worcester Road, Rexdale, Ontario

IMPORTANT NEW

READINGS IN EDUCATIONAL PSYCHOLOGY

by George J. Mouly,
University of Miami

Carefully abridged articles, accompanied by editorial commentary, are included for their significance to educational psychology and for their wide coverage of the field. These selections introduce the type of primary data which underlie psychological generalizations and principles of importance to the educative and pedagogical enterprise and acquaint the student with the scholars whose ideas constitute the substance of educational psychology.

May 1971/576 pages/$5.95 (*tent.*)

ENCOUNTERS WITH THE SELF

by Don E. Hamachek,
Michigan State University

The author synthesizes and integrates some of the basic propositions of self-concept theory along with a discussion of some of the major contributions to the understanding of self-concept in behavior. Short personal anecdotes, brief case history selections, and short counseling and therapy excerpts are used throughout the book as illustrative materials.

January 1971/264 pages/$4.95 paper

EDUCATIONAL RESEARCH: READINGS IN FOCUS

by Irving J. Lehman
and William A. Mehrens,
both of Michigan State University

Assembled primarily for students taking research courses in education, this collection of readings provides a discussion of the principles to be learned in the critiquing of research. The book is divided into five major types of research: historical, descriptive, correlational, causal-comparative, and experimental.

March 1971/512 pages/$6.95 paper (*tent.*)

TRAINING CREATIVE THINKING

by Gary A. Davis, University of Wisconsin and Joseph A. Scott, Chico State College

This volume brings together for the first time an up-to-date collection of scholarly readings from psychology, education, and industry that pertain to training creative problem-solving. Together the topics in the book present a comprehensive picture of *what* to teach when one teaches creativity, *how* to teach it, and *why* it should be taught.

January 1971/256 pages/$4.95 paper (*tent.*)

HUMAN DEVELOPMENT AND COGNITIVE PROCESSES

by John Eliot, University of Maryland

This book provides the reader with a broad frame of reference to an involved and abstract field of inquiry. It is divided into four parts and organized into at least six, sequentially arranged essays based upon four of Israel Scheffler's epistemological questions about knowledge. Part One considers cognitive development within the larger context of human development. Part Two centers upon perception and perceptual development. Part Three covers language and language acquisition. Part Four considers thinking and the elements of thought. Each part contains a list of alternate readings of varied complexity and focus.

January 1971/480 pages/$7.00 (*tent.*)

BOOKS FROM HRW

CONTEMPORARY INFLUENCES IN EARLY CHILDHOOD EDUCATION

by Ellis Evans,

University of Washington

The author covers pre-primary educational programs and the broad educational strategies applicable to three- through six-year-old children. He approaches the subject in terms of principles of learning and development reflected in these programs, the philosophical, psychological, and cultural issues associated with contemporary early childhood education, and the importance of evaluation at this stage.

January 1971/320 pages/$6.95 (tent.)

IMAGERY AND VERBAL BEHAVIOR

Allan Paivio,

University of Western Ontario

Table of Contents: Introduction. Development and Functions of the Symbolic Systems. Meaning and the Symbolic Process. Perception and the Symbolic Processes: Effects of Meaning. Perception and the Experimental Arousal of the Symbolic Process. Learning and Memory: Classical Mnemomic Systems. Stimulus Attributes and Memory. Meaning and Associative Learning. Verbal Mediation in Learning and Memory. Imagery Mediation in Learning and Memory. Distinguishing Imaginal and Verbal Mediators. Language and the Symbolic Processes: Linguistic Models and Associationism. Imagery and Language. Individual Differences in Symbolic Habits and Skills. Extensions and Speculations.

In preparation/704 pages/$12.95 (tent.)

Holt, Rinehart and Winston, Inc.

383 Madison Avenue, New York 10017

THE SEXUAL LIFE OF MAN

Herant Katchadourian

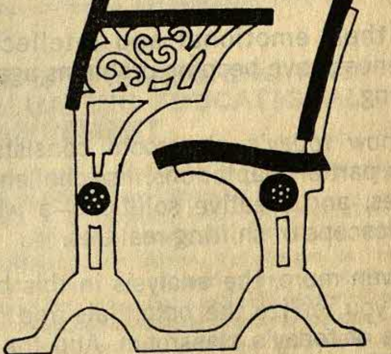
and Donald T. Lunde,

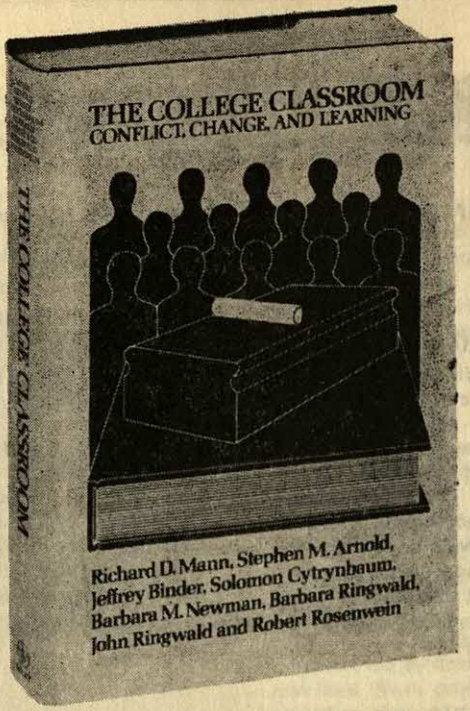
Stanford University Medical School

Table of Contents: Human Sexuality: Fundamental Questions. Anatomy of Sexual Organs. Physiology of Sexual Functions. Sexual Hormones. Conception and Pregnancy. Contraception. Psychosexual Development. Sexual Behavior. Auto-Eroticism. Sexual Intercourse. Variations and Deviations of Sexual Behavior. Sexual Disorders. The Erotic in Art. The Erotic in Literature. The Erotic in Film. Sex and the Law. Morality and Sex.

In preparation/512 pages/$10.00 (tent.)

Write to Mrs. Hope Ludlow if you would like a Holt traveler to visit you and tell you more about these books. Please mention this ad when writing.




THE COLLEGE CLASSROOM
CONFLICT, CHANGE, AND LEARNING
THE COLLEGE CLASSROOM
Richard D. Mann, Stephen M. Arnold,
Jeffrey Binder, Solomon Cytrynbaum,
Barbara M. Newman, Barbara Ringwald,
John Ringwald and Robert Rosenwein

Can Students Learn in Today's College Classroom?

Crumbling methods of education . . . student unrest . . . mammoth college administrations. Yet, learning must go on.

How?

Through the interpersonal dynamics of the classroom.

That's what *The College Classroom* is all about. It shows how yesterday's material-covering method leaves today's students cold.

How their emotional and intellectual awareness have become problems against learning.

And how today's classroom consists of crises, partial adaptations, new challenges, failures, and creative solutions—a whole kaleidoscope of shifting realities.

But even more, the analysis in this book helps you realize the potentials and difficulties of today's classroom. And the necessity for creating a new supportive teaching culture. One that welcomes innovation and bolsters self-awareness.

The new roads to learning that will work for you.

THE COLLEGE CLASSROOM: Conflict, Change, and Learning. By RICHARD MANN, STEPHEN M. ARNOLD, JEFFREY L. BINDER, all of the University of Michigan; SOLOMON CYTRYNBAUM, Yale University; BARBARA M. NEWMAN, BARBARA E. RINGWALD, JOHN W. RINGWALD, all of the University of Michigan; and ROBERT ROSENWEIN, Eastern Michigan University.

1970 389 pages $9.95

For more information, contact your local Wiley representative or write to Wayne Anderson, Promotion Manager.

wiley

JOHN WILEY & SONS, Inc.
605 Third Avenue
New York, N.Y. 10016
In Canada:
22 Worcester Road, Rexdale, Ontario

ed. psych.

Allyn and Bacon, Inc.

CONTEMPORARY ISSUES IN EDUCATIONAL PSYCHOLOGY

Harvey F. Clarizio, Robert C. Craig, and William A. Mehrens, all of Michigan State University
Contains thirty-eight pro and con issues dealing with meaningful controversies related to current educational practice. "This is the most up-to-date comprehensive compendium in Educational Psychology I have seen..." L. J. Mather, Howard University. 1970 6 x 9 Paperbound 747 pp.

NEW 1971 EVALUATING PUPIL GROWTH, Fourth Edition
Principles of Tests and Measurements

J. Stanley Ahmann, Colorado State University and Marvin D. Glock, Cornell University
This new text includes a balanced treatment of teacher made and standardized tests; a thorough treatment of all three taxonomies of educational objectives; new material on testing the disadvantaged; discussion of measuring personal-social adjustment; new material on the treatment of methods of measuring aptitudes and mastery criterion reference and norm reference concepts. 1971 6¼ x 9¼ Est. 608 pp.

NEW 1971 STATISTICS AND MEASUREMENT:
A Programmed Introduction, Second Edition

Lowell A. Schoer, University of Iowa
This programmed text, now in its second edition, presents statistical concepts necessary for dealing with norms, predictions, validity, reliability, regression and the standard error of measurement in a brief, concise manner. No statistical or mathematical background is needed. 1971 8½ x 11 Paperbound Est. 272 pp.

HUMAN DEVELOPMENT IN WESTERN CULTURE,
Third Edition

Harold W. Bernard, Oregon State System of Higher Education
The third edition of this outstanding text has been completely up-dated to include information on the most recent research and experimental programs in the study of human development. Suitable for courses in human development, educational psychology, and child development, the text now includes a chapter on education for the disadvantaged. 1970 6½ x 9. 642 pp.

NEW 1971! MEASURING AND EVALUATING EDUCATIONAL ACHIEVEMENT

J. Stanley Ahmann, Colorado State University; and Marvin D. Glock, Cornell University
This text is an abridged version of EVALUATING PUPIL GROWTH: PRINCIPLES OF TESTS AND MEASUREMENT, 4th Edition, and focuses on the construction and use of testing and non-testing instruments for measuring pupil achievement. Contains a complete treatment of teacher-made tests, a thorough treatment of all three taxonomies of educational objectives and mastery criterion reference and norm reference concepts. 1971 6¼ x 9¼.

ALLYN AND BACON, INC. COLLEGE DIVISION DEPT. 893
470 Atlantic Avenue, Boston, Mass. 02210

PEOPLE WATCHING

Curriculum and techniques for exploring the behavioral sciences in the classroom.

Editor, Barbara Ellis Long

Presents articles, special features, new techniques, reviews, and programs, which deal with aspects of the behavioral sciences and their application in the school curriculum. **People Watching** will provide a vehicle for the exchange of information among teachers, curriculum developers, research investigators, and community mental health personnel, on the topic of behavioral science curricula and applied programs in elementary and secondary schools.

A major section of **People Watching** will be concerned with information and ideas for the teacher. Articles by teachers will allow for a sharing of ideas and experiences.

VOLUME I, NO. 1, SPRING 1971, Bi-annual

ANNUAL SUBSCRIPTION RATES: INDIVIDUALS $5.00

INSTITUTIONS $10.00

CHILD CARE QUARTERLY

Editor, Jerome Beker

Child Care Quarterly: An Independent Journal of Day and Residential Child Care Practice, will begin publication early in 1971. *Child Care Quarterly* has been conceived as the first independent professional publication dedicated specifically to the improvement of child care practice in a variety of settings and to the advancement of this rapidly growing field. Designed to serve child care workers, their supervisors, and other personnel in child care settings as well as instructors in the field of child care, it will provide a channel of communication and debate including articles on practice, selection and training, theory and research, and professional issues. Relevant book reviews, news notes, and other features will also be included.

ANNUAL SUBSCRIPTION RATES: INDIVIDUALS $10.00

INSTITUTIONS $20.00

TWO, FOUR, SIX, EIGHT, WHEN YOU GONNA INTEGRATE?

By Frank A. Petroni, Ph.D., Ernest A. Hirsch, Ph.D. and C. Lillian Petroni

The social-psychological impact of attending an integrated high school is examined in a case studies approach using student subgroups ranging from Black "elites" to White racists.

1971; APPROX. 270 pgs. HARDBOUND, $9.95

Illustrated Books For Children 4 to 8 on Psychologically Relevant Themes

I HAVE FEELINGS

By Terry Berger Photographed by I. Howard Spivak

I HAVE FEELINGS covers seventeen different feelings, both good and bad, and the situations that precipitated each one. Each feeling is presented by a situation, the feeling that results and finally by an explanation of that feeling.

HARDBOUND, $3.95

THE MAN OF THE HOUSE

By Joan Fassler HARDBOUND, $3.95

ALL ALONE WITH DADDY

HARDBOUND, $3.95

ONE LITTLE GIRL

HARDBOUND, $3.95

Behavioral Publications

2852 Broadway-Morningside Heights

New York, New York 10025

BASIC
BOOKS ARE
MORE THAN
TIMELY

BASIC STUDIES ON READING

Harry Levin & Joanna P. Williams, Editors

Reading is a complicated process about which we know too little. From this assumption, the distinguished contributors to this book proceed to three main areas of basic study—linguistics and reading, psychological processes, and social variables in reading research. Many of these studies emerged from Project Literacy, a five-year endeavor carried out under the auspices of the U.S. Office of Education. $15.00

THE EXCEPTIONAL CHILD ANNUAL, 1969

Jerome Hellmuth, Editor

The special problems and special needs of the exceptional child are explored in this selection of the most important writings in the field. Among the areas covered are: mental retardation, brain damage, perceptual-motor problems, learning disability, reading disorders, communication disorders, the disadvantaged child, and other topics of current concern. $15.00

READING DISABILITY

Diagnosis and Treatment

Second Revised Edition

Florence G. Roswell & Gladys Natchez

When the complex process of reading falters, the causes for the disability are equally complex. Drawing on their experience in the classroom and as clinical psychologists, the authors examine the casual factors of reading failure and set forth methods of diagnosis and treatment. This new edition is enriched by the inclusion of important references and materials that have appeared since the first edition. $6.95

FAMILY PROCESS

Nathan W. Ackerman, Editor

The leading pioneer of family therapy presents here some of the best and most influential work that has been done in the field of the family. Culled from studies that have appeared over the last nine years in the journal *Family Process*, this selection of articles examines all aspects of the family—the nature of family living, the growth of a science of the family, and the methods for treatment of troubled families. $10.00

BASIC BOOKS, Inc.
404 Park Avenue South
New York, N.Y. 10016

DEVELOPMENTAL PSYCHOLOGY

Editor: Boyd R. McCandless

Developmental Psychology contains articles which represent the broad range of growth and development and their major associated variables. Chronological age as well as sex, socioeconomic status and effects of physical growth variables (including time of maturation and body build) are all considered relevant developmental variables.

Although the adolescent and the aging population appear to deserve particular attention at this time, *Developmental Psychology* indeed covers the entire span. Cross-species articles and articles concerning developmental research with retardates are also included.

The articles in *Developmental Psychology* are not limited to any one approach such as "experiments versus studies" or "cognition versus motivation".

Bimonthly. Two volumes per year.
Subscription Price: $20.00 per year

AMERICAN PSYCHOLOGICAL ASSOCIATION
1200 Seventeenth Street, N. W.
Washington, D.C. 20036

THE STUDY OF GAMES
By ELLIOTT MORTON AVEDON and BRIAN SUTTON-SMITH, both of Teachers College, Columbia University

This text presents a multi-disciplinary approach to the understanding of the structure and function of games and their role as instrumental factors in civilizations throughout history.

It covers the origins, history, use, structure, and function of each game. 1971 448 pages $9.95

CHILD'S PLAY
Collected readings on the biology, ecology, psychology and sociology of play
By ROBIN HERRON, Texas Medical Center, Houston; and BRIAN SUTTON-SMITH, Teachers College, Columbia University

This collection of readings deals with the history of child's play over the past 50 years. It provides an interdisciplinary review of the subject and contains a lot of unpublished relevant material.
1971 416 pages $9.95

SOCIAL DEVELOPMENT AND PERSONALITY
Edited by GEORGE G. THOMPSON, Ohio State University
With the collaboration of **Francis J. Di Vesta** and **John E. Horrocks,** both of Ohio State University
Foreword by **Merl C. Wittrock**

This text of readings deals with the social development of male and female adolescents. The papers demonstrate how social actions, social preferences, social motivations, social learnings, and the internalization of social controls appear and shape the individual's social development during childhood and adolescence.

They also provide new information on theory, methodology, and substantive findings.

Extensive references are included.

A volume of Readings in Educational Research Series
1971 Approx. 1024 pages In Press

INTELLECTUAL DEVELOPMENT
By PAULINE S. SEARS

This text consists of a number of readings on the development of intelligence and intellectual processes.

It gives a brief historical introduction to the measurement of intelligence (IQ tests); discusses the child's generalized concepts of objects, actions, and relationships in the environment (Piaget's theory); and shows how language influences the child's thinking.

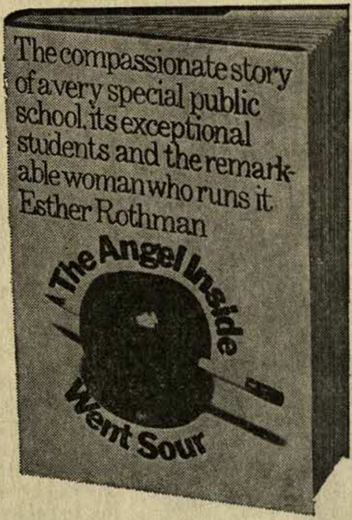
A volume of Readings in Educational Research Series
1971 Approx. 512 pages In Press

PEOPLE IN CONTEXT:
Measuring Person-Environment Consequence in Education and Industry
By GEORGE G. STERN, Syracuse University

This text shows how to use the Activities Index, the College Characteristics Index and other related instruments to measure the psychological characteristics and compatability of people and their environments. It clarifies college cultures, analyzes representative institutions in detail, and develops the theoretical rationale of behavioral interaction. 1970 402 pages $13.95

wiley
JOHN WILEY & SONS, Inc.
605 Third Avenue, New York, N.Y. 10016
In Canada: 22 Worcester Road, Rexdale, Ontario

Written by an extraordinary teacher with the warmth and humor of *Up the Down Staircase* and the dramatic impact of *Death at an Early Age.*


The compassionate story of a very special public school, its exceptional students and the remarkable woman who runs it
Esther Rothman
The Angel Inside Went Sour

The Angel Inside Went Sour

by ESTHER ROTHMAN

Only a remarkable teacher like Dr. Esther Rothman could have written this compassionate, volatile, bluntly honest account of ten years spent running The Livingston School, New York's only public day high school devoted to teenage girls other schools could not handle. Dr. Rothman explains the unconventional (but workable) techniques she uses to reach these disturbed and hostile girls by telling the stories of the teenagers who attend Livingston, their many problems, and the unusual answers that have been found to help them. Her book reveals frankly the violence, obscene language, sexual problems, neglect, and poverty that hallmark the lives of the girls, yet it bursts with the warm humor that makes Esther Rothman able to carry on no matter what the day brings.

At all bookstores $7.95

McKAY

JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 62 April 1971 Number 2

D[illegible] No 360
6·9·71
Lil

CONTENTS

Multidimensional Analysis Procedures for Measuring Self-Concept in Poverty Area Classrooms 95
DAN LANDIS, JOHN L. HAYMAN, JR., AND WILLIAM S. HALL

Student Evaluations of Teacher Behaviors as Estimations of Real-Ideal Discrepancies: A Critique of Teacher Rating Methods 104
CHARLES F. LEVINTHAL, LEONARD M. LANSKY, AND O. ERNEST ANDREWS

Team Teaching Compared with Traditional Instruction in Grades Kindergarten through Six 110
FEN RHODES

Stimulus Complexity and Intramodal Reaction Time in Retarded Readers 117
MARION BLANK, THOMAS J. HIGGINS, AND WAGNER H. BRIDGER

Individual Differences in Visual and Auditory Memory 123
ARTHUR R. JENSEN

Simultaneous and Successive Discrimination of Similar Letters 132
JOANNA P. WILLIAMS AND MARGARET D. ACKERMAN

Word Scanning Rate for Good and Poor Readers 138
LEONARD KATZ AND DAVID A. WICKLUND

Order Effects in the Affective Learning Styles of Overachievers and Underachievers 141
JOSEPH F. RYCHLAK AND THOMAS J. TOBIN

Feedback Procedures in Programmed Instruction 148
RICHARD C. ANDERSON, RAYMOND W. KULHAVY, AND THOMAS ANDRE

Selection and Socialization: A Study of Engineering Student Attrition 157
ROBERT ATHANASIOU

Transfer from Teaching to Learning 167
K. KILBY LONG

© 1971 by the American Psychological Association, Inc.

Journal of Educational Psychology
1971, Vol. 62, No. 2, 95–103

MULTIDIMENSIONAL ANALYSIS PROCEDURES FOR MEASURING SELF-CONCEPT IN POVERTY AREA CLASSROOMS[1]†

DAN LANDIS
Educational Testing Service
Princeton, New Jersey

JOHN L. HAYMAN, JR.
Council of the Great City Schools
Washington, D.C.

AND WILLIAM S. HALL[2]
Princeton University

The present study is an initial attempt to apply the individual differences multidimensional analysis procedure of Tucker and Messick to the description of self-image. Four hundred eighty-seven fourth-grade subjects in 16 poverty area classrooms were administered an instrument in which 12 critical concepts were presented by the method of complete triads. After interconcept distance estimates were obtained, the data were subjected to the Tucker and Messick procedures. Three factors were obtained, representing subgroups of the 16 classrooms. Within each group, the number of concept dimensions varied from three to four. Tentative interpretations of each concept dimension were made but will require other experimental verification before the results can be accepted. The implications and feasibility of the method in large scale evaluation programs are discussed.

In recent years sizable sums of money have been spent on compensatory education programs, with the major portion of these funds being expended under Title I of the Elementary and Secondary Education Act (ESEA) of 1965 (PL 89-10). A commonly used objective for these programs is the "improvement of the self-image of the 'disadvantaged' child." Unfortunately, there has been little agreement on the nature of self-image let alone on measurement methods.

While most of the evaluations of such programs have tended to ignore this particular program objective, those taking it seriously have typically employed variations of the Semantic Differential (Osgood, Suci, & Tannenbaum, 1957) in assessing self-image. These evaluations have produced less than impressive findings. Failure to find convincing results can be related to at least two reasons. First, there is the possibility that the programs have little or no effect on the child's self-image and second, the method typically used may not be sufficiently sensitive to measure changes in self-image over short periods such as those characteristic of compensatory intervention program. Nevertheless, change in self-image will continue to be a critical goal for intervention programs.

Complete evaluations and modifications of these programs will require a method capable of assessing self-image that can be used reliably and easily on a relatively large scale. Such a method should not be concerned with producing individual data but rather with ascertaining reliable data on the structure of self as it might vary across programmatic units, for example, classes, schools, districts, programs, etc. The present article describes a first attempt to develop a useable method for the assessment of self-image at this level.

Prior research on the nature of, and methodology for measuring, self-image suggests three broad conclusions: (*a*) most of the available measures require individual rather than group administration of all items; (*b*) researchers have been more concerned with evaluative aspects of the self-image while giving relatively little attention to a simple structural description; and

---

[1] Requests for reprints should be sent to Dan Landis, Educational Testing Service, Princeton, New Jersey.

[2] The authors wish to express their appreciation to Clifton E. Mayfield and Ellery Pearson for their support and invaluable aid in the execution of this study and to Robert Linn and Rae M. Landis for thoughtful criticisms of earlier drafts of this paper. They should also like to thank David R. Saunders for performing the data analysis.

Copyright © 1971 by the American Psychological Association, Inc.

(*c*) when the self is described in terms of one's relationship to significant others, the "others" are almost from within the family constellation as understood from a middle-class perspective.

With regard to the first conclusion, one notes that with the exception of research using the Semantic Differential, most attempts to study self-image have been based on individually rather than group-administered instruments (e.g., Koslin, Amarel, & Ames, 1969; Long & Henderson, 1968). Since many program characteristics interact in an uncontrolled and unknown fashion, evaluation typically requires a large sample. In this case, cost considerations would mitigate against the use of individually administered instruments for evaluation purposes. This does not mean that individual administration is not a good method but only that it may be infeasible for evaluation in contrast to reseach purposes.

The second conclusion, an apparent overconcern with the evaluative[3] rather than the descriptive aspects of self-image, is perhaps a reflection of the recent concern with the "disadvantaged" segment of our society. Thus, several studies have attempted to get at the matter of self-image among disadvantaged children by comparing them with children from more advantaged backgrounds. Findings from these investigations are contradictory, for instance, Soares and Soares (1969) found that disadvantaged children did not necessarily have negative self-images. On the other hand, Long and Henderson (1968) found that they did. Additionally, Long and Henderson (1968) report that disadvantaged children identify less with their fathers, have lower self-esteem, and less realistic perceptions of their own skin color. These contradictory results indicate that evaluative comparisons among groups are premature and that methodological improvements are needed. The suspicion remains, for example, that making an evaluative judgment based on a comparison of privileged and nonprivileged subjects employing the linguistic conventions of the former group yields a distorted picture. (Baratz & Baratz, 1970).[4] For this reason the interpretation of these studies must remain uncertain until we have a better structural picture of the self which is set in an appropriate subcultural milieu.

Attempts to relate the self to "significant others" are not only theoretically important (Mead, 1934) but should also go a long way toward making sense out of such terms as "negative self-image." Examples of studies embodying this approach are those of Schwartz (1966) and Ludwig and Maehr (1967). Schwartz (1966) found that children's ego functioning was significantly related to parents' perception of the child which in turn was related to kind of parent-child interaction. Further evidence of the role of the "significant other" in influencing self-concept can be seen in the work of Ludwig and Maehr (1967) who found that the self-concept of seventh- and eighth-grade boys changed as a function of the reaction of significant others. These and similar studies do, however, suffer from an important theoretical defect. The significant others that are selected are those important to the middle-class child. For a ghetto child with a profoundly different family structure (Billingsley, 1969; Schulz, 1969) there are much more important "others" to whom he relates himself, for instance, the peer group, teachers, etc. Certainly an adequate picture of the self must indicate the relationship between the child's view of himself and his view of those individuals, institutions, and concepts that are critically important for his development.

All of the studies described above, while interesting from the standpoint of analyzing individual behavior, have not been designed to deal sufficiently with either the social or the educational milieu and their relationship to self-image. Because of this failure, they have questionable relevance for use in evaluating the effect of social and educational intervention programs. It seems clear that the self-concept must be viewed in relationship to other components of functioning and their buttresses. This very statement suggests that multidimensional analysis may be a useful approach. Furthermore, if our aim is to compare children in programmatic units, then we should look at the differences between these units in the multidimensional structuring of self. The individual differences multidimensional analysis method of Tucker and Messick (1963) would seem to be ideally suited for making these contrasts. While applications of this method, (e.g., Landis & Slivka, 1970; Landis, Silver, Jones, & Messick, 1967; Silver, Landis, & Messick 1966) have looked at differences between individuals, there is no reason why the unit of analysis should not be a particular meaningful group rather than the individual.

---

[3] This discussion does *not* mean that we dismiss the evaluative component of self-concept. Rather we are concerned with (*a*) the *value* terms that the researcher places on the end points of the evaluative factor (e.g., Coopersmith, 1967, p. 26–27) and (*b*) the reference group used to define a "healthy" self-image.

[4] Osgood (1967) has pointed to similar problems which occur when one attempts to do cross-national research. He points out that one must be extremely careful to make sure that the stimuli used maintain constant referents even after translation. In engaging in research using black subjects we must recognize the existence of a distinct language structure with different linguistic and affective properties.

One further condition here is that the instrument must require a judgment that is not highly dependent on reading skill if it is to be used en masse with young children. This condition automatically rules out the Semantic Differential as a viable instrument. Schwartz and Tangeri (1965) used the Semantic Differential with sixth-grade delinquent boys and found that the only way that they could obtain reliable data was to use no more than 10 scales with four concepts each and explain verbally and repetitiously each scale throughout the administration. The present study was designed to measure self-concept in a way that would be useful for the evaluation of large scale intervention programs while, at the same time, theoretically avoiding most of the problems discussed above.

METHOD

Subjects

The subjects were 487 fourth-grade children attending 16 classes[5] distributed in five schools throughout the School District of Philadelphia. Four out of five schools were generally classified as serving poverty areas for the purposes of receiving Title I funds. The fifth school was not so classified and indeed had a relatively small proportion of children from poverty backgrounds attending the school. Table 1 provides the breakdown of the numbers of children in each of the 16 classrooms participating in this study, as well as the racial composition of each of the five schools.

TABLE 1

POPULATION OF CLASSROOMS IN THE STUDY TOGETHER WITH RACIAL COMPOSITION OF EACH SCHOOL

School	Classroom	Number of children	Racial composition[a]	
			White	Negro
A			887	881
	1	36		
	2	32		
	3	37		
	4	35		
	5	30		
B			429	258
	1	25		
	2	26		
C			1,120	1,120
	1	32		
	2	34		
	3	35		
	4	37		
	5	26		
D			430	44
	1	16		
	2	30		
E			307	217
	1	36		
	2	20		

[a] Data are for the total school. Precise data at the classroom level are not available. However, given the urban characteristic of the school systems it can be assumed that the ratio of Negro to white children will be higher at the fourth grade than for the total school.

The Instrument—Most Like Questionnaire (MLQ)

The following 12 concepts were selected as being important for the educational and social definition of the child's self-image. (*a*) teachers, (*b*) policemen, (*c*) school principals, (*d*) school counselors, (*e*) social workers, (*f*) soldiers, (*g*) Negroes, (*h*) white people, (*i*) Puerto Ricans, (*j*) my friends, (*k*) I, (*l*) school. Each concept was combined with each of the other concepts in a set of complete triads and presented in the following form: Teachers are most like (A) soldiers or (B) Puerto Ricans.

Since the number of complete triads that can be derived from 12 concepts is very large (660), it was felt that it would be difficult, if not impossible, to present all such triads to each pupil. Furthermore, in view of the main goals of the instrument, such data would not be crucial. Thus, six forms were constructed, each having 110 randomly selected triads. These forms were randomly intermixed in a set for each classroom and handed out in this manner. Consequently, each classroom received approximately equal numbers of each of the forms and all triads were used in each classroom.

---

[5] All schools had mean reading retardation of 1½ to 3 years when grade level was used as the point of reference (Mayfield & Landis, 1966).

Analysis of Data

The data were analyzed first to determine whether major response (or concept meaning) categories, which would differentiate among the separate classroom units, existed, and second, if there were such categories, to identify them and their characteristics. As implied, complete classrooms were the basic unit of analysis.

To get at the problem of determining the existence of distinct categories, a method was needed to compute and analyze the degree of similarity in total concept structure between each pair of classrooms. This was accomplished in two stages. First, interstimulus distance estimates among the 12 concepts were obtained for each classroom directly from the pupils' similarity judgments, following the method outlined by Torgerson (1958). This produced 16 12 × 12 matrices—one for each classroom—showing distance estimates between each concept. These matrices were then subjected to a Tucker-Messick Individual Differences Multidimensional Analysis (Tucker & Messick, 1963), and the outcome was a 16 × 16 matrix, with all entries showing similarity in total concept structure between classrooms. The 1,5 cell of this matrix, in other words, indicated the degree of similarity in total concept structure between Classrooms 1 and 5. This "similarity in concept structure" matrix was then factor analyzed, and the resulting factors defined distinct classroom groups that structured the 12 original stimuli in essentially similar fashion. Each factor or category might be referred to as a particular point of view. Among the 16 classes, three significant factors, or viewpoints, were recovered. These three accounted for approximately 95% of the variance in the similarity judgments.

TABLE 2

Rotated (Equamax) Factor Loadings of 16 Classrooms on Three Points of View

Classroom	Points of view		
	A	B	C
B-2	.5517	.6711	.6002
C-1	.8788	.6837	.8240
C-2	.9890	.7883	.8621
B-1	.9425	1.2536	1.0024
A-5	.7513	.6973	.8444
E-2	.8534	.7849	.8254
C-3	1.2017	1.0706	.9765
D-1	1.1921	1.2181	.9770
E-1	.6937	.8493	.6429
A-1	.7176	.5996	.6214
C-4	1.4634	1.1155	.8698
C-5	.7202	.8471	1.2268
A-3	1.2450	1.0814	.8584
D-2	.9915	1.4808	.7495
A-4	1.0991	1.0468	.7566
A-2	.5208	.5482	.8268

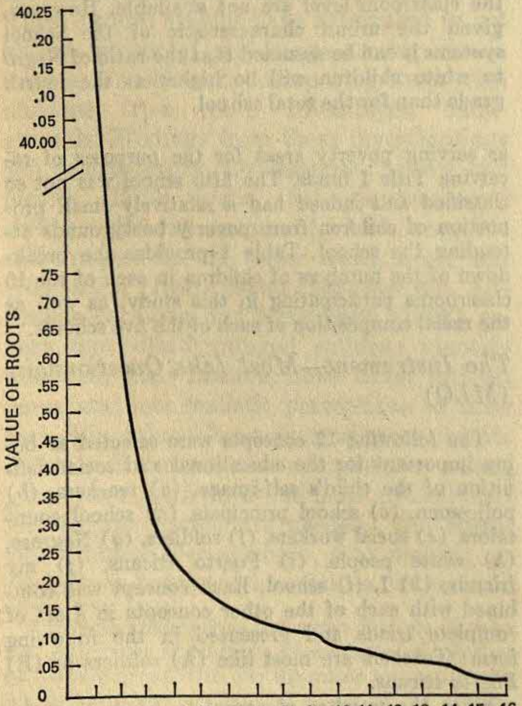


Fig. 1 Characteristic root plot.

Figure 1 shows the characteristic root plot obtained from this analysis. The break in the curve after the third root indicates that none of the other factors is significant.

To discover the structural characteristics of each category or point of view, classrooms were grouped according to which of the three points of view the factor loadings indicated they belonged. Returning to concept loadings in the Tucker-Messick analysis, the 12 stimuli were separately scaled for each of the three classroom groups. This amounted to a kind of factor analysis of concepts for each group, and the resulting factors indicated the concept structure for the particular point of view under consideration. For two of the three viewpoints, the 12 concepts could be scaled parsimoniously along three dimensions; four dimensions were required for the third point of view.

RESULTS

Table 2 presents the rotated (Equamax; Saunders, 1962) factor loadings for the 16 classrooms on the three viewpoints. Inspection of Table 2 indicates that Viewpoint A is most clearly marked by Classroom C-4, Viewpoint B by Classroom D-2, and Viewpoint C by Classroom C-5. Interestingly enough, Classrooms C-4 and C-5 are both located in the same building of a poverty area school with approximately equal numbers of black and white children (See Table 1). On the other hand, Classroom D-2 is located in a school with very few black children.

Figures 2, 3, and 4, present the stimulus dimensions for each point of view. By noting where each concept falls along each factor, some hint of structure of the concept space for the children in those classes in a given point of view may be given.

Factor A-I may represent either the view of white students in a largely Negro school, or that of Negroes rejecting their own skin color. The relationship of school and social figures to "policeman" is interesting and confirms popular reports.

Factor A-2 is more clearly seen to be a racial factor. Here the self-concept, concept of Negroes, concept of Puerto Ricans, are

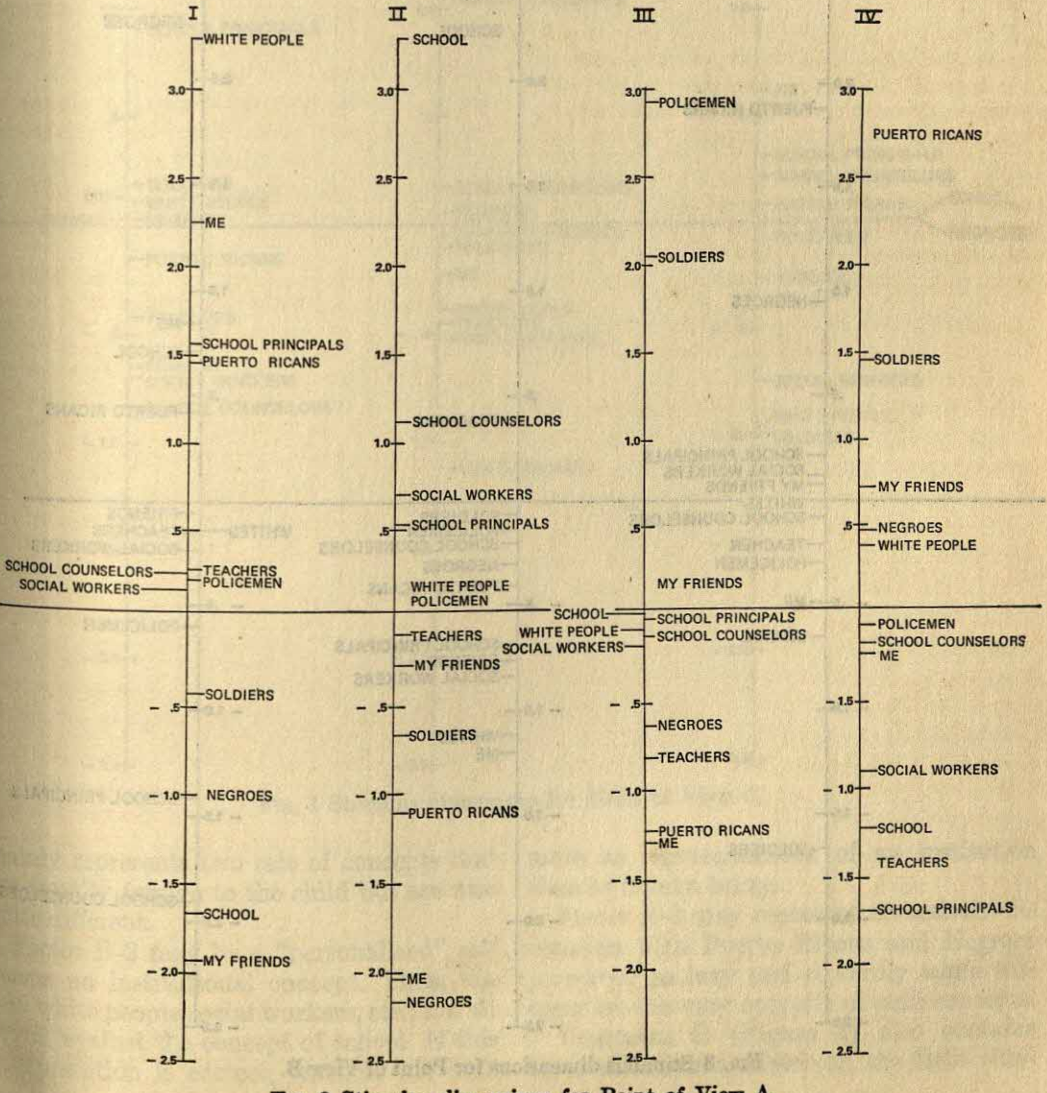


FIG. 2 Stimulus dimensions for Point of View A.

set off against the concept of school. One might argue that what is being represented here is the distance of the school from those individuals that are relevant to the lives of the children.

Factor A-3 is perhaps a potency-impotency dimension with policeman and soldiers at one end and self and racial grouping at the other end.

Factor A-4 seems to be a school versus nonschool dimension. Thus, principals, teachers, school, and social worker are all at the same end.

Viewpoint B consists of only three dimensions. Notably lacking here is the potency dimension found in Viewpoint A.

Factor B-1 may be an educated versus noneducated dimension. Or perhaps it

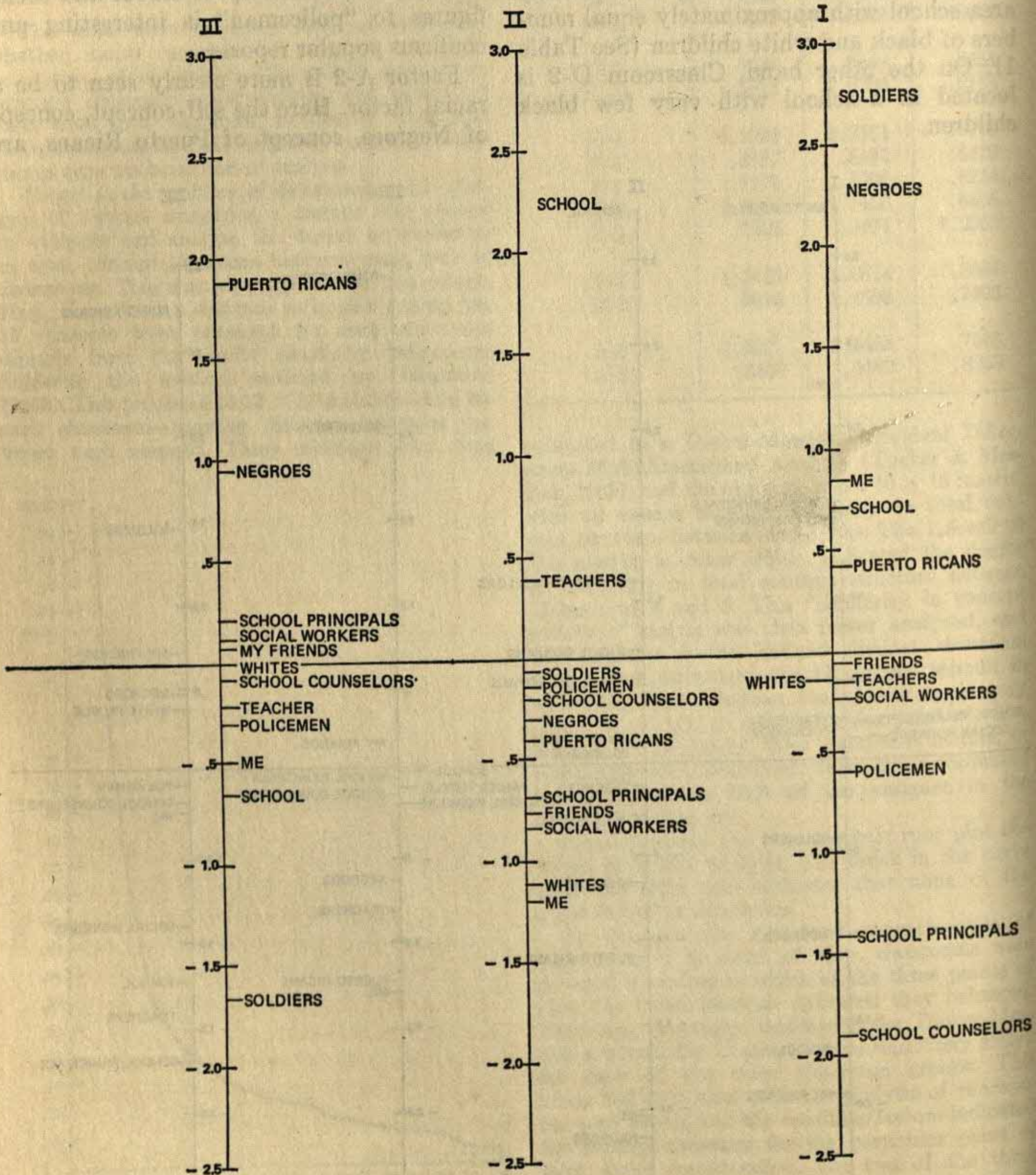


FIG. 3 Stimulus dimensions for Point of View B.

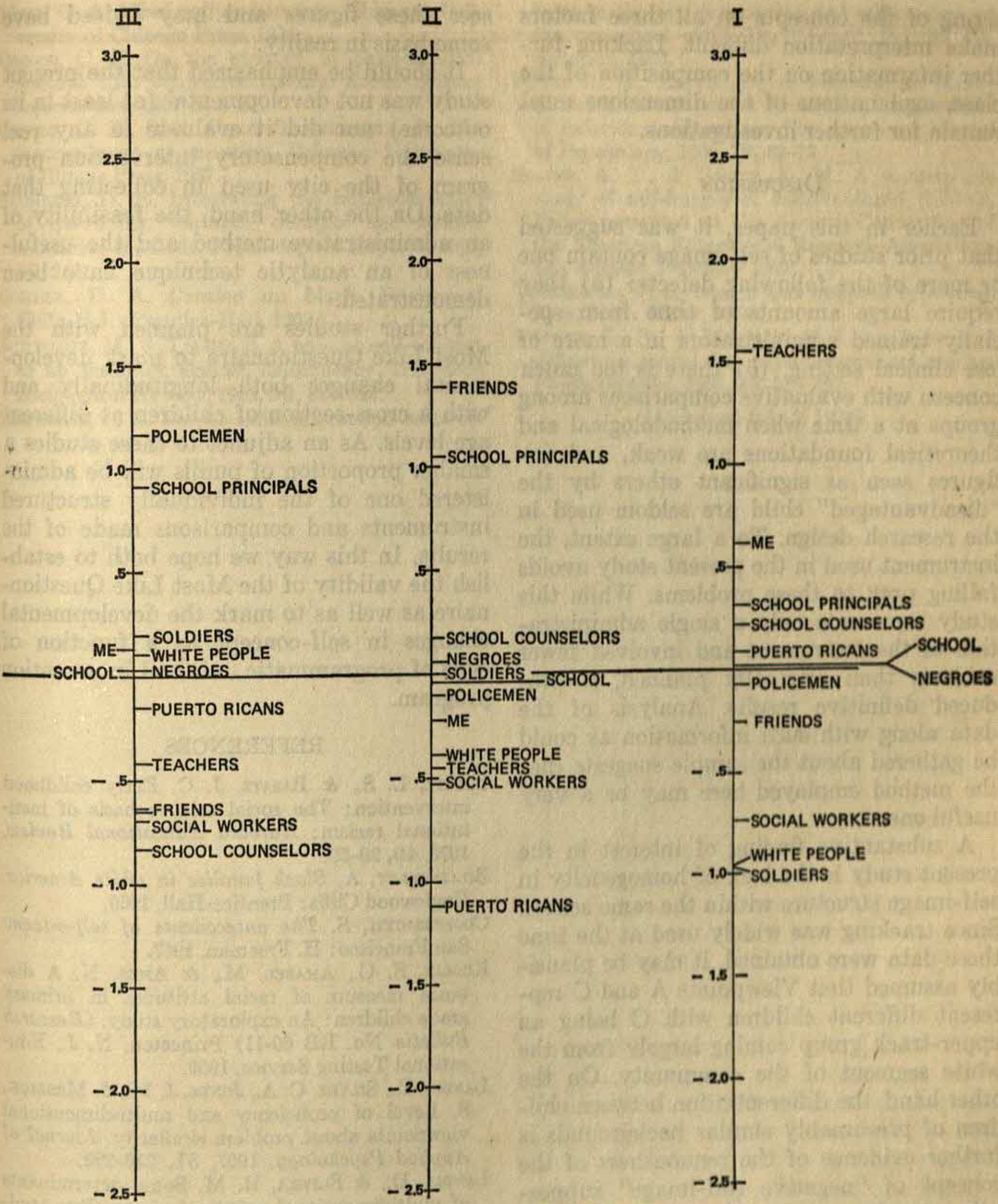


FIG. 4 Stimulus dimensions for Point of View C.

mainly represents two sets of concepts that are equally foreign to the child but are also quite different.

Factor B-2 may be a "personalized" self versus an institutional concept. Thus, the self, white people social workers, etc., are all set off against the concept of school. If this interpretation is correct, teachers are seen more as representations of an institution than as human beings.

Factor B-3 may represent an activity dimension with Puerto Ricans and Negroes perceived as lazy and cowardly while soldiers are the very opposite of such concepts.

Viewpoint C (Figure 4) also contains three dimensions. However, the tight clustering of the concepts on all three factors make interpretation difficult. Lacking further information on the composition of the class, explanations of the dimensions must remain for further investigations.

Discussion

Earlier in this paper, it was suggested that prior studies of self-image contain one or more of the following defects: (*a*) they require large amounts of time from specially trained administrators in a more or less clinical setting, (*b*) there is too much concern with evaluative comparisons among groups at a time when methodological and theoretical foundations are weak, and (*c*) figures seen as significant others by the "disadvantaged" child are seldom used in the research design. To a large extent, the instrument used in the present study avoids falling prey to these problems. While this study represents only a single administration of the instrument and involved fewer subjects than originally planned, it produced definitive results. Analysis of the data along with such information as could be gathered about the sample suggests that the method employed here may be a very useful one.

A substantive finding of interest in the present study is the lack of homogeneity in self-image structure within the same school. Since tracking was widely used at the time these data were obtained, it may be plausibly assumed that Viewpoints A and C represent different children with C being an upper-track group coming largely from the white segment of the community. On the other hand, the differentiation between children of presumably similar backgrounds is further evidence of the tenuousness of the concept of "negative self-image" supposedly prevalent among ghetto children. We have made no evaluative judgment on these structures—they have been presented as found. It seems quite reasonable that each structure has certain survival value and represents the optimum for the given environment. From a middle-class perspective we may not think it "healthy" to identify teachers with policemen; but the fact that the child does this is a reflection of how he sees these figures and may indeed have some basis in reality.

It should be emphasized that the present study was not developmental (at least in its outcome) nor did it evaluate in any real sense the compensatory intervention program of the city used in collecting that data. On the other hand, the feasibility of an administrative method and the usefulness of an analytic technique have been demonstrated.

Further studies are planned with the Most Like Questionnaire to mark developmental changes both longitudinally and with a cross-section of children at different age levels. As an adjunct to these studies a smaller proportion of pupils will be administered one of the individually structured instruments and comparisons made of the results. In this way we hope both to establish the validity of the Most Like Questionnaire as well as to mark the developmental changes in self-concept as a function of type of programmatic unit and intervention program.

REFERENCES

Baratz, L. S., & Baratz, J. C. Early childhood intervention: The social science basis of institutional racism: *Harvard Educational Review,* 1970, **40,** 29–50.

Billingsley, A. *Black families in white America.* Englewood Cliffs: Prentice-Hall, 1969.

Coopersmith, S. *The antecedents of self-esteem.* San Francisco: H. Freeman, 1967.

Koslin, S. C., Amarel, M., & Ames, N. A distance measure of racial attitudes in primary grade children: An exploratory study. (*Research Bulletin* No. RB 69-11) Princeton, N. J., Educational Testing Service, 1969.

Landis, D., Silver, C. A., Jones, J. M., & Messick, S. Level of proficiency and multi-dimensional viewpoints about problem similarity. *Journal of Applied Psychology,* 1967, **51,** 216–222.

Landis, D., & Slivka, R. M. Some determinants of multidimensional points of views: Two studies. *Perceptual and Motor Skills,* 1970, **31,** 659–670.

Long, B. H., & Henderson, E. H. Self-social concepts of disadvantaged school beginners. *Journal of Genetic Psychology,* 1968, **113,** 41–51.

Ludwig, D. J., & Maehr, M. L. Changes in self-concepts and stated behavioral preferences. *Child Development,* 1967, **38,** 453–67.

Mayfield, C. E., & Landis, D. *Evaluation of FY 1965 ESEA projects in Philadelphia.* (Franklin Institute Technical Report 1-178) Philadelphia: Franklin Institute, 1966.

MEAD, G. A. *Mind, self and society.* Chicago: University of Chicago Press, 1934.

OSGOOD, C. E. On the strategy of cross-national research into subjective culture. *Social Science Information,* 1967, **6,** 5–37.

OSGOOD, C. E., SUCI, G., & TANNENBAUM, P. *The measurement of meaning.* Urbana: University of Illinois Press, 1957.

SAUNDERS, D. R. Integrating the implementation of quartimax, varimax, oblimax and related rotational methods. *Psychological Reports,* 1962, **10,** 241.

SCHULZ, D. A. *Coming up black.* Englewood Cliffs, N.J.: Prentice-Hall, 1969.

SCHWARTZ, M., & TANGRI, S. A note on self-concept as an insulator against delinquency. *American Sociological Review,* 1965, **30,** 922–926.

SCHWARTZ, S. C. Parent-child interaction as it relates to the ego functioning and self-concept of the pre-school child. Unpublished doctoral dissertation, 1966.

SILVER, C. A., LANDIS, D., & MESSICK, S. Multidimensional analysis of visual form: An analysis of individual differences. *The American Journal of Psychology,* 1966, **79,** 62–72.

SOARES, A. T., & SOARES, L. M. A comparative study of self-images of disadvantaged children. Paper presented at the Annual Convention of the American Educational Research Association, Los Angeles, February 1969.

TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

TUCKER, L. R., & MESSICK, S. J. An individual differences model for multi-dimensional scaling. *Psychometrika,* 1963, **38,** 333–367.

(Received July 9, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 2, 104–109

STUDENT EVALUATIONS OF TEACHER BEHAVIORS AS ESTIMATIONS OF REAL-IDEAL DISCREPANCIES:

A CRITIQUE OF TEACHER RATING METHODS[1]

CHARLES F. LEVINTHAL
University of Michigan

LEONARD M. LANSKY[2]
University of Cincinnati

AND O. ERNEST ANDREWS
State University of New York at Stony Brook

Ratings imply comparative judgments between the values of the observer and his observations, for instance, a student's ratings of his teacher are estimates of the discrepancy between the student's ideals for the teacher's behavior and what he sees the teacher do. However, most methods for collecting teacher ratings make assumptions about ideals and about the discrepancies between ideals and observed behavior. In order to assess the relevance of direct measurement of ideals, students were asked to rate a teacher *and* to report their ideals for the teacher's behaviors. The judgments of ideal behavior varied across students and items; there were also interactions between ideals and observed responses. Thus, the authors recommend new approaches to teacher ratings.

Students have ideas about ideal teacher behaviors. Students also see their teachers behave. One way to interpret students' evaluations of teachers is to say that each student estimates the discrepancy between his ideal for that behavior and what he sees his instructor do. When the instructor's behavior comes close to the student's ideal, the teacher receives a high rating; if the teacher is far from the student's ideal, the teacher receives a low rating.

Most approaches to teacher evaluation do not assess the student's ideal directly. For example, students may be asked to rate the teacher's "flexibility" on a 5-point scale ranging from "outstanding" to "poor." In this approach, the student makes an explicit evaluation (i.e., estimates a discrepancy) based upon an implicit observation. The response of "poor" for a given attribute would be interpreted as the maximum discrepancy between the ideal and observed behavior for each student. However, the extent to which similar low ratings among a set of students truly reflects the constancy of the actual psychological discrepancies depends upon the assumption that all students have the same ideal point along the attribute dimension.

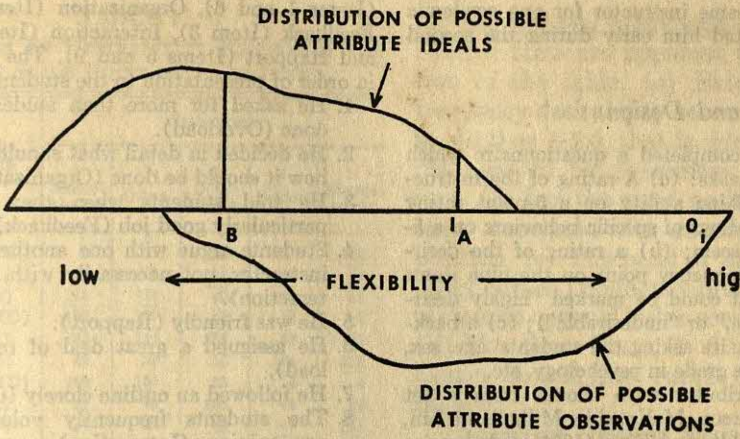
Coombs (1964) has discussed the problem of interpreting such evaluations. He has suggested that students stretch their maximum discrepancy distances to fit their evaluation scales. Figure 1 contains a hypothetical distribution of ideal and observed responses on a rating scale for "flexibility." If, for example, Student A has an ideal point at I_A at the middle of the distribution of possible observations, the maximum discrepancy can be only about half as great as that of Student B, having his ideal point at I_B, far to the left end of the observation distribution. A rating of "poor" by Student B would represent a much greater discrepancy between his ideal and his observation than a rating of "poor" by Student A, even though both students made the same rating response. According to this interpretation, individual evaluations on an outstanding to poor scale are basically incomparable and cannot be legitimately combined when the value systems of the raters differ from one another.

In another approach to teacher ratings, the student is presented with a specific pattern of behavior, such as "He was flexible." Students then describe their instructor in

---

[1] This paper is an expanded version of one presented at the annual meeting of the Midwestern Psychological Association in Chicago, 1969.

[2] Requests for reprints should be sent to Leonard M. Lansky, Department of Psychology, McMicken Hall, University of Cincinnati, Cincinnati, Ohio 45221.


DISTRIBUTION OF POSSIBLE ATTRIBUTE IDEALS
I_B
I_A
o_i
low
FLEXIBILITY
high
DISTRIBUTION OF POSSIBLE ATTRIBUTE OBSERVATIONS

$$|o_i - I_A| < |o_i - I_B|$$

FIG. 1. Theoretical distributions of ideal and observed scores on the attribute "flexibility" (Adapted from C. H. Coombs, 1964).

terms of the frequency of the behavior on a scale from "almost always" to "almost never." Here, the student is making an explicit observation which implies the evaluation. Perhaps it is more accurate to say that the experimenter infers the evaluation. For example, the intuitive evaluative inference of the observation that the instructor performed some positive teaching behavior "almost always" is that the instructor was "outstanding" in this attribute. The intuitive inference, however, may not accurately reflect the valuative system of the student from whom the observation was obtained, much less the interaction of that system with the observation itself.

Although both types of rating scale appear in Werdell's (1966) summary of teacher rating scales for the National Student Association, he does not discuss the relevance of students' ideals for teacher behaviors; nor do the scales take account of this idea. So far as we can tell, there are no published reports on college teaching which assess students' ideals directly. There is, however, a report to the Office of Education (Rickfelder, Brown, & Milholland, 1968) of a study in which students completed ideal and actual *Q* sorts about college teachers' behaviors.

The experimental literature on the effect of feeding ratings back to elementary school teachers (Gage, Runkel, & Chatterjee, 1963) and other public school personnel (Daw & Gage, 1967) contains several examples of the use of ideal and actual ratings. In these studies, shifts toward or away from the ideals are the criteria for the effect of the feedback. Similar studies on college teachers and courses are lacking.

The present study was designed to explore the effects of direct assessment of students' ideals and their observations of a teacher. Two specific questions were asked:

1. How would the ideals be distributed for each teacher behavior? A preliminary question was, of course, could ideals be reliably assessed?

2. What relationships would obtain between the students' ideals and the observations of teacher behaviors?

The prediction was that ideal behaviors and the relationships between ideal and observed behaviors would vary, indicating that new approaches to teacher ratings should be seriously considered.

METHOD

Subjects

The subjects were 263 University of Cincinnati students, 143 males and 120 females, enrolled in the year-long introductory psychology class. They had all had the same instructor for one academic quarter; they rated him early during the second quarter.

Questionnaire and Design

Each student completed a questionnaire which contained three tasks: (*a*) A rating of the instructor's overall teaching ability on a 5-point rating scale, and nine ratings of specific behaviors on a 5-point frequency scale; (*b*) a rating of the desirability of each frequency point on the nine items (Each scale point could be marked "highly desirable," "acceptable," or "undesirable"); (*c*) a background questionnaire asking the students' age, sex, previous quarter's grade in psychology, etc.

The nine attributes were chosen from a set described by Isaacson, McKeachie, Milholland, Lin, Hofeller, Baerwaldt, and Zinn (1964) as being independent of general teaching ability according to a factor analysis. The factors were Overload (Items 1 and 6), Organization (Items 2 and 7), Feedback (Item 3), Interaction (Items 4 and 8), and Rapport (Items 5 and 9). The actual items, in order of presentation to the students, were:

1. He asked for more than students could get done (Overload).
2. He decided in detail what should be done and how it should be done (Organization).
3. He told students when they had done a particularly good job (Feedback).
4. Students argue with one another or with the instructor, not necessarily with hostility (Interaction).
5. He was friendly (Rapport).
6. He assigned a great deal of reading (Overload).
7. He followed an outline closely (Organization).
8. The students frequently volunteered their own opinions (Interaction).
9. He was permissive and flexible (Rapport).

Each item was presented in the form below which also indicates one set of possible responses:

1	2	[✓ 3]	4	5
Almost always	Often	Occasionally	Seldom	Almost never
Undesirable	Undesirable	Acceptable	Highly desirable	Acceptable

Two forms of the questionnaire were used. For one sequence, the form was as described above. The student performed the second task by turning back to the nine observation scales and making his desirability responses in the five spaces provided beneath each scale point. For another group, the first two tasks were reversed, that is, the student estimated the desirability of each frequency point on the nine scales first, then he gave the instructor an overall rating and turned back to the nine scales to record his observations on the nine attributes.

Two different testing sessions were held in order to fit students' schedules. At each session, approximately half of those tested were randomly assigned to one of the two questionnaire sequences.

Results and Discussion

There were no statistically significant differences between testing sessions, questionnaire sequences, or between the two sexes on any of the measures. Thus, all subjects were combined for further analyses. The comparisons of the two sessions constituted a reliability check on the measures.

The first issue was the pattern of the desirability responses. It was expected that there would be considerable spread in the responses for a given behavior and that different behaviors would have different desirability patterns.

Table 1 contains the raw desirability score distributions for each point on each item. The modal desirability score for each scale point is underlined. Inspection of the patterns reveals considerable spread within each item and variability across items.

However, in order to look more closely at the data, scores of $+1$, 0, and -1 were given, respectively, to desirability responses of "highly desirable," "acceptable," and "undesirable." There were two assumptions behind this scoring scheme: (*a*) the "acceptable" response represented a relatively noncommittal attitude; and (*b*) the psychological distances between adjacent responses on the 3-point scale were approximately equivalent.

The next step was to compute two group desirability scores for each item. The first score (frequency desirability) consisted of the average desirability for each frequency point. Thus, the range of frequency desirability was from -1.0 to $+1.0$.

The second score (optimal value) focused on the "highly desirable" response. For each item, each student's optimal value was defined as the mean of his "highly desirable" response(es) on the frequency scale. Thus, any student who rated only one point as

TABLE 1

FREQUENCIES OF DESIRABILITY RESPONSES FOR EACH SCALE POINT ON EACH ITEM

Item no.	Desirability score	Scale point				
		1 Almost always	2 Often	3 Occasionally	4 Seldom	5 Almost never
1	+1 (HD)	9	17	53	87	110
	0 (A)	23	72	168	141	72
	−1 (U)	231	174	42	35	81
2	+1 (HD)	129	127	47	11	8
	0 (A)	85	114	178	78	35
	−1 (U)	49	22	38	174	220
3	+1 (HD)	160	165	55	4	4
	0 (A)	67	78	196	94	25
	−1 (U)	36	20	12	165	234
4	+1 (HD)	49	76	102	23	27
	0 (A)	57	105	143	161	63
	−1 (U)	157	82	18	79	173
5	+1 (HD)	208	157	28	8	6
	0 (A)	45	98	210	47	9
	−1 (U)	10	8	25	208	248
6	+1 (HD)	9	34	74	57	45
	0 (A)	80	133	166	123	63
	−1 (U)	174	96	23	83	155
7	+1 (HD)	135	129	31	9	10
	0 (A)	77	111	194	77	30
	−1 (U)	51	23	38	177	223
8	+1 (HD)	127	148	74	10	16
	0 (A)	80	91	185	125	28
	−1 (U)	56	24	4	128	225
9	+1 (HD)	73	102	81	12	9
	0 (A)	91	123	168	113	29
	−1 (U)	99	38	14	138	225

Note.—$N = 263$. The modal desirability score for each scale point is underlined. Abbreviated: HD = highly desirable, A = acceptable, U = unacceptable.

"highly desirable" on an item received that frequency point as his optimal value for that item. The median optimal value for all students was then computed for each item.

Table 2 contains the frequency desirability and optimal value scores for each item. Several facts are apparent from an inspection of the table. (*a*) Except for Item 5, frequency desirability scores are never close to +1.0 or −1.0, that is, ideal responses do not tend to be unanimous. (*b*) Only one item (5) has the most positive desirability score associated with an extreme frequency position. (*c*) The peak of the group desirability curves (frequency desirability scores) is close to optimal value scores for all items. (*d*) For three items, 4, 6, and 9, the frequency desirability curves are inverted **U** shaped.

The second question focused on the interaction between students' ideals and their actual ratings of the teacher behaviors. In order to examine this question, we first selected only those students who had marked only one point on the frequency scale as "highly desirable." Since the "highly desirable" response could fall on any point for each item, there were potentially five different groups of students: those with the "highly desirable" response located in the "almost always" position (HD----), those with "highly desirable" in the "often" position (-HD---), etc. We then looked at the actual ratings of students in each of the five groups. If, for a given item, the students' observations of their teacher (the frequency ratings) were indeed independent of their optimal values, then the five groups would show similar distributions of such ratings; the ideal scores would be irrelevant to the observed scores. However, if there were dif-

TABLE 2

MEAN FREQUENCY DESIRABILITY AND MEDIAN OPTIMAL VALUE FOR EACH ITEM

Item no.	Factor	Frequency desirability M					Optimal value
		1 Almost always	2 Often	3 Occasionally	4 Almost never	5 Never	
1	Overload	−.84	−.60	.04	.20	.11	4.4
2	Organization	.30	.40	.03	−.62	−.81	1.5
3	Feedback	.47	.55	.16	−.61	−.87	1.5
4	Interaction	−.41	−.02	.32	−.21	−.56	3.0
5	Rapport	.75	.57	.01	−.76	−.92	1.5
6	Overload	−.63	−.24	.19	−.10	−.42	3.5
7	Organization	.32	.40	−.03	−.64	−.81	1.5
8	Interaction	.27	.47	.27	−.45	−.79	1.5
9	Rapport	−.10	.24	.25	−.48	−.82	2.0

TABLE 3

DISTRIBUTION OF FREQUENCY RATINGS FOR FIVE OPTIMAL VALUE GROUPS FOR ITEMS 2, 3, AND 9 SEPARATELY

Item	Optimal value pattern	Number of subjects who rated teacher					
		1 Almost always	2 Often	3 Occasionally	4 Seldom	5 Almost never	Total
2	HD----	11	3	1	4	4	23
	-HD---	11	7	1	2	—	11
	--HD--	3	5	7	5	2	22
	---HD-	1	2	1	—	—	4
	----HD	—	—	—	1	2	3
	Total	26	17	10	12	8	63
3	HD----	8	3	2	—	1	14
	-HD---	2	4	4	1	2	13
	--HD--	—	2	8	3	2	15
	---HD-	—	—	—	1	—	1
	----HD	1	—	—	1	—	2
	Total	11	9	14	6	5	45
9	HD----	1	1	5	1	—	8
	-HD---	—	4	5	3	—	12
	--HD--	—	3	14	10	2	29
	---HD-	—	2	—	—	—	2
	----HD	—	—	—	—	1	1
	Total	1	10	24	14	3	52

Note.—Abbreviated: HD = highly desirable.

ferent interaction patterns, then the question of directly assessing students' ideals warrants further study.

Table 3 contains the data for three items, 2, 3, and 9. The Index of Predictive Association (Goodman & Kruskal, 1954) for Items 2 and 3 is .26; for Item 9, it is .04. Thus, it appears that for Items 2 and 3 valuative systems and frequency observations interact whereas for Item 9, these two variables seem to be independent.

Insofar as this was an exploratory study, the results are only suggestive. We used only one teacher, one set of students, and one set of teacher behaviors. In addition, the behavioral and desirability ratings were made on the same page of the questionnaire; the subjects could see the one as they did the other. The proximity in space and time of the ideal and observed responses might have produced spuriously high interactions. However, the variability among the three items in Table 3 poses a new problem: differential susceptibility to this procedural artifact. Yet another methodological issue concerns the methods for collecting the behavioral and desirability ratings themselves. In a study closely related to the one reported here, Rickfelder et al. (1968) used *Q* sorts on the behavior of the real teacher and ideal teacher to assess the similarity of the scaling of the attributes in the two procedures. Other methods are, of course, possible. Despite the limitations cited above, the present findings support the view that direct assessments of students' values are relevant for understanding teacher ratings.

There are other reasons to pursue this approach: the current interest in values and value conflict in education; the concern for developing new approaches for the college classroom; the interest in providing students and teachers more useful feedback; and the need in several areas of psychology for more complex models for measuring complex social interactions. The availability of college classrooms and the growth of teacher rating programs offers an opportunity for further testing the idea of teacher evaluation as an estimation of discrepancies between ideals and observed behaviors.

REFERENCES

COOMBS, C. H. *A theory of data.* New York: Wiley, 1964.

DAW, R. L., & GAGE, N. L. Effect of feedback from teachers to principals. *Journal of Educational Psychology,* 1967, **58,** 181–188.

Gage, N. L., Runkel, P. J., & Chatterjee, B. B. Changing teacher behavior through feedback from pupils: An application of equilibrium theory. In W. W. Charters & N. L. Gage, *Readings in the social psychology of education.* Boston: Allyn & Bacon, 1963.

Goodman, L. A., & Kruskal, W. H. Measures of association for cross-classifications. *Journal of the American Statistical Association,* 1954, **49,** 732–764. Cited by W. L. Hays, *Statistics for psychologists.* New York: Holt, 1963.

Isaacson, R. L., McKeachie, W. J., Milholland, J. E., Lin, Y. G., Hofeller, M., Baerwaldt, J. W., & Zinn, K. L. Dimensions of student evalutions of teaching. *Journal of Educational Psychology,* 1964, **55,** 344–351.

Rickfelder, A., Brown, B., & Milholland, J. E. The teacher *Q*-sort. MIM Project. (Report to the Office of Education, IV-6) Washington, D. C.: United States Government Printing Office 1968.

Werdell, P. *Course and teacher evaluation* (2nd ed.) Washington, D. C.: United States National Student Association, 1966.

(Received December 22, 1969)

Journal of Educational Psychology
1971, Vol. 62, No. 2, 110-116

TEAM TEACHING COMPARED WITH TRADITIONAL INSTRUCTION IN GRADES KINDERGARTEN THROUGH SIX

FEN RHODES[1]
California State College at Long Beach

During the 1968-69 school year the effectiveness of an elementary school with team teaching at all grade levels (K-6) was compared against a control school with traditional, self-contained classes. Pupil achievement in reading, spelling, and arithmetic, and pupil attitude toward school and learning were measured at the beginning and end of the year. Attitudes of teachers and parents at the year end were also assessed. In no instance was team teaching found to be superior to traditional classroom instruction, and it was significantly worse with respect to average reading gain ($p < .01$) and change in pupil attitude ($p < .05$). Teachers of team-taught classes were more positive ($p < .05$) in their job attitude than teachers at the control school. Parents of children attending the team versus control school were found to be equally favorable in their opinion of school effectiveness ($p > .05$).

Team teaching is a relatively recent innovation in American education. According to Anderson (1964), " . . . the first full-scale project in team teaching was launched in an elementary school in Lexington, Massachusetts . . . in the 1957-58 school year [p. 127]." In 1964 a survey of school districts in seven western states found 30% with team-teaching programs (Borg & Brite, 1967). Most of these programs were quite new, having been in operation 1 year or less, and few were more than 2 years old.

Since then team teaching has increased rapidly in popularity. Unfortunately, the publication of definitive research evaluating the effects of team teaching has not kept pace. There is certainly no scarcity of reports describing and assessing various team-teaching projects throughout the country, but the number of evaluations employing control groups and providing adequate statistical treatment of data is quite small. Typically, such studies have compared team with traditional teaching at a single grade level. The investigations of Knox (1966) at Grade 1, Burningham (1968) at Grade 4, and Georgiades and Bjelke (1966) at Grade 9 are illustrative. Composition and functioning of instructional teams differ from study to study, and there is little consistency in terms of dependent variables measured.

In spite of these differences the majority of controlled evaluations are agreed on the following points:

1. Achievement of pupils under team instruction is no better than achievement in traditional self-contained classes.
2. Pupil attitude is the same under team and regular instruction.
3. Teacher attitude is more positive under team instruction.

Those studies the author has seen that have reached substantively different conclusions are open to question on methodological grounds or are subject to criticism with respect to interpretation of data. Three cases in point are Sharken (1963), who ignored nonsignificant t tests and based his conclusions concerning achievement on the number of observations in the expected direction; Georgiades and Bjelke (1966), who did not control for differences in initial achievement level in addition to IQ; and Zweibelson, Bahnmuller, & Lyman (1965), who in their analyses of variance interpreted the teaching-mode main effect rather than its nonsignificant interaction with the pre-post effect as representing differential changes in pupil attitude.

---

[1] Requests for reprints should be sent to Fen Rhodes, Psychology Department, California State College at Long Beach, Long Beach, California 90801.

OBJECTIVES OF PRESENT STUDY

This study compared team teaching with traditional instruction at all elementary grades (K–6), and was designed to answer the following questions:

1. Do pupils in the elementary grades learn more under team instruction?
2. Do children who are team-taught have a more favorable attitude toward school and learning than that of pupils enrolled in regular classes?
3. How does the attitude of team teachers toward their job compare with the attitude expressed by teachers of self-contained classes?
4. How does the attitude of parents toward a team-teaching school compare with the attitude of parents whose children are in a traditional school?

It must be emphasized that the answers to these questions provided by the present study are capable of generalization to other team-teaching situations only insofar as what is called "team teaching" elsewhere bears a resemblance to the instructional program described in this report.

DESCRIPTION OF TEAM AND CONTROL SCHOOLS

The two schools compared in this study were part of an elementary school district serving a suburban city in the greater Los Angeles area. At the time of the study the team program was in its second full year of operation. Every effort was made to select a control school that matched the team school as nearly as possible with respect to total enrollment, pupil-teacher ratio, pupil turnover, parent socioeconomic status, tenure of principal, and age of physical plant.

Classrooms at the team-taught school had been designed to accommodate team instruction. For Grades 1–6 each room was L-shaped and had an area three times that of a self-contained classroom. Kindergarten was taught in a rectangular classroom twice the usual size. Rooms were carpeted to reduce noise, and the building was air-conditioned.

A separate temporary structure at the team-taught school housed six traditional classrooms in which self-contained classes were taught—one at Grades 1 and 4, and two at Grades 2 and 3. Assignment of pupils to the self-contained classes was not random, and tended to exclude children whose achievement fell markedly above or below grade level. These classrooms, like those at the control school, were not carpeted or air-conditioned.

Instructional teams in Grades 1–6 were composed of three teachers, the kindergarten team of two. In each team one teacher was designated as leader. The team leader coordinated planning of the overall academic program for the class, and arranged for whole-group lectures or demonstrations by individual team members. She also coordinated a number of flexible groupings for reading and arithmetic instruction. Except for these types of activities pertaining to all class members, pupils were seated in separate groups of 30 and instructed by one team member, who had primary academic and record-keeping responsibility for these students. Team leaders also served as grade-level coordinators, exercising informal supervision over curriculum and instructional methods for the self-contained classes.

At both team and control schools fifth and sixth graders spent one-half of each day learning basic skills subjects. The other half was devoted to foreign language, social studies, science, art, music, and physical education. These subjects were taught by pairs of resource teachers, supplemented by district-wide instructional television.

Two teacher aides were employed at the team school, one serving Grades K–3 and the other Grades 4–6. These aides performed a variety of clerical tasks for the team teachers and also assisted them in developing special curriculum materials. The control school had one teacher aide to assist the resource teachers in Grades 5 and 6.

The team school had a very active parent organization, whose members devoted much effort to the preparation of classroom materials requested by teachers. In certain cases parents were regularly present in the classroom to assist teachers. The control school also had an active parent group, but it did not appear to be so deeply or enthusiastically involved in day-to-day activities.

METHOD

Selection of Sample

One-fourth of the pupils in each classroom of the team and control schools were selected at random for the comparison study. Pupils who moved or who changed classrooms for any reason during the year were eliminated from the sample, as were any children enrolled in special-education programs. Out of the original group a total of 316 pupils remained at the end of the school year. This represented 138 pupils from the control school and 147 pupils from team-taught classes. An additional 31 pupils included in the sample were enrolled in self-contained classes operated at the team school.

Evaluation of Achievement

Achievement in reading, spelling, and arithmetic was measured by the Wide Range Achievement Test (WRAT), 1965 revision. This test was administered at both schools in October 1968 and June 1969. Testing was conducted by classroom teachers, who had received prior training in administration of the WRAT.

IQ Measurement

IQ data for pupils in the study were collected at various times during the school year. Intelligence was measured by the Goodenough-Harris Drawing Test (Grades K–3, 5), the Otis-Lennon Mental Ability Test (Grade 4), and the California Test of Mental Maturity (Grade 6).

Measurement of Pupil Attitude

The favorability of each child's attitude toward school and learning was assessed by his teacher in October and in June. The instrument used was a 45-item weighted checklist containing descriptive examples of various pupil behaviors reflecting different degrees of enthusiasm toward school and learning.[2] Individual attitude scores were computed by averaging the weights associated with those statements checked for each pupil.

Statements for the attitude checklist were selected and scaled using the Thurstone method of equal-appearing intervals. During the school year preceding the comparison study, behavioral statements were solicited from teachers in Grades K–6, representing both team-taught and self-contained classes. A pool of 215 nonredundant statements describing a wide variety of pupil behaviors potentially reflective of favorability of attitude toward school and learning was developed for judging.

These statements were ordered in nine categories from unfavorable to favorable by 15 teachers from Grades K–3 and 12 from Grades 4–6. Twenty-one of the teachers were from team-taught and six from self-contained classes. Since none of the statements described behaviors peculiar to either the team or self-contained setting, this imbalance was not considered to bias the scaling.

Judges were instructed to categorize statements solely with regard to indicated attitude toward school and learning, and were specifically cautioned against rating particular behaviors high or low based on the level of academic achievement they might imply. The degree of independence achieved is demonstrated by the correlation between rated pupil attitude and achievement test scores at year end, which did not exceed .03 for the total sample of 316 children.

To insure that behavioral statements included in the final checklist reflected similar pupil attitudes at both upper and lower grade levels, the median rating of each statement by teachers from Grades K–3 was compared with the median rating by Grade 4–6 teachers. Those statements showing a substantial difference in scale value were eliminated from further consideration. For the 45 statements in the completed checklist there was a .36 average absolute difference in judged scale value between upper and lower grades. The average interquartile range of pooled judgments for these 45 statements was 1.3. Statements for the checklist were selected to provide a rectangular distribution of scale values across the nine categories.

Measurement of Teacher and Parent Attitudes

Attitude of teachers toward their job and the attitude of parents toward the school and their child's classroom experience were assessed at the end of the year by means of Likert-type opinion questionnaires. In neither questionnaire was any direct reference made to team teaching or teaching methods. Rather, statements were phrased in terms of situational specifics reflecting important goals and relevant concerns of teachers or parents that obtain under any type of instruction. Respondents expressed their opinion regarding each statement on a six-category scale from strongly agree (+3) to strongly disagree (−3). Neutral (0) responses were not permitted.

Questionnaires (coded to preserve respondents' anonymity) were sent to each teacher and to the parents of each child included in the study. Correlations between individual statement responses and total score were obtained for each questionnaire, and statements with *r*s below .20 were eliminated. Four items were deleted from the teacher scale (leaving 28), and one from the parent scale (leaving 29). Questionnaires were then rescored on the basis of items remaining.

Analysis of Data

Teaching Mode × Grade Level analyses of variance and covariance were employed to evaluate pupil achievement and attitude, as well as teacher and parent attitudes, under team and regular instruction. One-way analyses were used to compare teacher and parent attitudes in team-taught versus self-contained classes within the team school. Teaching mode and grade level were taken as fixed factors in making *F* tests; subjects were considered random.

---

[2] Copies of the behavior checklist are available from the author upon request.

Because there were unequal numbers of observations under the various combinations of teaching mode and grade level, an unweighted-means analysis technique was used in all cases (Winer, 1962). Variance analyses were performed using a computer program essentially identical to Veldman's (1967) AVAR23. The same program was used for the covariance analyses, with individual scores adjusted by linear regression to partial out the effect of the covariate.

RESULTS

Table 1 shows the mean achievement gains of pupils in reading, spelling, and arithmetic under team compared with traditional instruction, as well as the relative changes in pupil attitude that occurred during the school year. Each of the means listed is an unweighted average of the individual means obtained across all grade levels for that teaching mode. In no case was team teaching superior to traditional classroom instruction, and it was significantly worse with respect to average reading gain and, to some extent, degree of improvement in pupil attitude.

TABLE 1

MEAN IMPROVEMENT IN ACADEMIC ACHIEVEMENT AND PUPIL ATTITUDE UNDER TEAM VERSUS TRADITIONAL INSTRUCTION

Area	Grades K–6[a]			Grades 1–4[b]		
	Team	Non-team	Diff.	Team	Non-team	Diff.
Reading[c]	14.2	17.5	−3.3**	16.4	19.3	−2.9**
Spelling[c]	11.5	13.6	−2.1	11.7	12.5	−0.8
Arithmetic[c]	11.6	13.3	−1.7	10.1	11.2	−1.1
Pupil attitude[d]	−0.3	0.0	−0.3*	−0.2	−0.1	−0.1

[a] Team versus control school—unadjusted means (excludes self-contained classes in team school).

[b] Team-taught versus self-contained classes within team school—adjusted means (except pupil attitude).

[c] Expressed in months of achievement (10 months = 1 school year).

[d] Expressed on 9-point scale reflecting favorability of attitude toward school and learning.

* $p < .05$.

** $p < .01$.

There was no significant difference in initial achievement level or attitude of pupils in the team and nonteam groups (Tables 2 and 3). Mean pupil IQ was identical (105) in the team and control schools, and within the team school did not differ significantly between self-contained and comparable team-taught classes (IQ = 102 versus 107). Achievement gains in the latter instance are shown analyzed with the effect of IQ partialed out to permit accurate comparison with gains made by the team and control schools. *F* values reported for the covariance analyses do not differ markedly from those obtained using unadjusted achievement scores.

The significant grade-level effect shown in Tables 2 and 3 for initial achievement and improvement in reading, spelling, and arithmetic reflects the fact that pupils in the upper grades received relatively higher WRAT raw scores, and—as normally occurs on this test—they realized smaller raw-score gains during the school year. The

TABLE 2

SUMMARY OF ANALYSES OF VARIANCE FOR PUPIL MEASURES: TEAM VERSUS CONTROL SCHOOL (GRADES K–6)

Source	*df*	*F*—Initial level					*F*—Improvement			
		IQ	Reading achievement	Spelling achievement	Arithmetic achievement	Attitude	Reading achievement	Spelling achievement	Arithmetic achievement	Attitude
Teaching mode (A)	1/271	.01	.72	1.08	3.08	.57	12.69**	1.76	1.23	4.59*
Grade level (B)	6/271	4.39**	248.74**	302.55**	349.69**	2.20	31.59**	23.85**	6.99**	2.63*
A × B	6/271	1.81	1.01	.89	1.58	1.35	2.42*	1.78	.38	3.79**

Note.—Self-contained classes in team school excluded from above analyses.

* $p < .05$.

** $p < .01$.

TABLE 3

SUMMARY OF ANALYSES OF VARIANCE AND COVARIANCE FOR PUPIL MEASURES WITHIN TEAM SCHOOL: TEAM-TAUGHT VERSUS SELF-CONTAINED CLASSES (GRADES 1–4)

Source	*df*	*F*—Initial level					*F*—Improvement			
		IQ	Reading achievement	Spelling achievement	Arithmetic achievement	Attitude	Reading achievement[a]	Spelling achievement[a]	Arithmetic achievement[a]	Attitude
Teaching mode (A)	1/94	3.80	1.81	1.72	.79	1.22	8.05**	.47	.98	.12
Grade level (B)	3/94	2.37	53.64**	72.19**	102.75**	.42	12.42**	6.30**	5.62**	1.79
A × B	3/94	1.69	.14	.15	.57	.06	2.22	1.80	1.92	2.36

[a] Covariance analysis with effect of IQ partialed out.
* $p < .05$.
** $p < .01$.

significant grade-level variation observed for IQ scores (Table 2) is likely a function of differences in the three tests used to measure intelligence.

The interaction between teaching mode (A) and grade level (B) for reading improvement in the team and control schools is graphed in Figure 1. It can be seen that this effect was caused largely by an absence in the third and fourth grades of superior improvement under nonteam instruction. Although there was no significant AB interaction for reading improvement within the team school (Table 3), it is noteworthy that only at the third-grade level did self-contained classes fail to improve substantially more than team classes. The AB interaction that occurred with respect to attitude improvement (Table 2) was significant beyond the .01 level. However, the absolute differences among cell means are so small that an explanation of their pattern would appear to be of little practical interest.

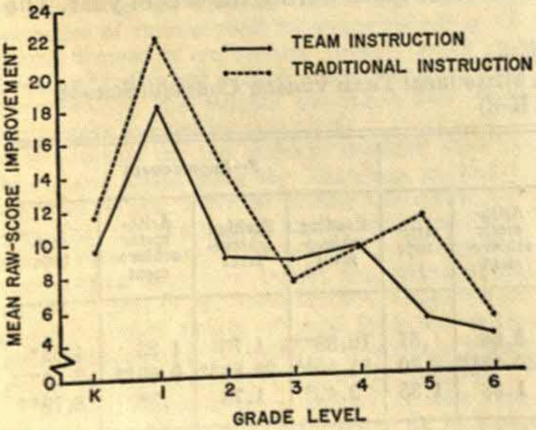


FIG. 1. Reading improvement by grade level under team and traditional instruction (team versus control school).

Concerning teacher and parent attitudes, it was found that teachers of team-taught classes were significantly more positive ($F = 5.68$, $df = 1/38$, $p < .05$) in their job attitude than teachers at the control school. Within the team school, however, there was no difference ($F = .00$, $df = 1/18$, $p > .05$) in the attitude of team teachers and teachers of self-contained classes. Parents of team-taught children were found to be no more or less favorable in their opinion of school effectiveness ($F = .00$, $df = 1/255$, $p > .05$) than were parents of children attending the control school. Similarly, no difference was found within the team school between parents associated with the two types of classes ($F = .86$, $df = 1/92$, $p > .05$). Ninety percent of the parents of children in the study responded to the attitude questionnaire.

DISCUSSION

The present findings parallel those of Boyles (1968) concerning team teaching in the first, seventh, and eighth grades. Using the Metropolitan Achievement Tests to measure academic progress and the California Test of Personality plus special rating scales to measure pupil adjustment, attitude, and self-concept, she found no significant difference in personality at the end of the school year between students in team-taught and those in self-contained classes.

However, there were significant achievement differences at every grade level, all favoring nonteam instruction. These occurred in the areas of reading (first and seventh grades), arithmetic (seventh grade), language (seventh and eighth grades), and social studies (seventh and eighth grades). Significant achievement differences were likewise obtained consistently in the present study, and all favored traditional instruction.

There was little change in pupil attitude during the school year under either type of instruction (Table 1). That the small change observed was in the direction of deterioration rather than improvement is not surprising in light of the finding of Flanders, Morrison, and Brode (1968) that "a significant loss in positive attitude of pupils toward their teachers and schoolwork [occurs] during the school year [p. 334]." What seems unusual is the fact that the observed deterioration in attitude was not greater under both teaching conditions. Perhaps this may be taken as evidence supporting the author's impression that rater bias was indeed present in this situation and, further, that it was reasonably balanced in intensity between the team and control schools.

It is emphasized that the nonsuperiority of team teaching and its failure in this instance to prove even equal to traditional instruction in every area cannot be attributed to any lack of enthusiasm on the part of the team teachers. Their job attitude was highly positive—and not just relative to that of teachers in the control school—as is indicated by their mean score of 5.8 on the 7-point attitude scale used. (The mean attitude score of control-school teachers was 5.2.) It is relevant that no teacher was forced into the team-teaching mold when the program was first set up; all were volunteers.

The WRAT, which was used to assess academic performance, is open to the usual criticism leveled against standardized achievement tests, namely, that they measure rather rote skills and knowledge with no emphasis on creativity or capacity for integration of information. Whether this be true or false, it remains that these rote skills are basic skills, ones that are necessary for adequate functioning in the areas to which they relate. That they are not alone sufficient does not diminish their essential importance as prerequisities to competence. Even if this and other achievement tests are as narrow in scope as their worst critics contend, it would still be reasonable to insist that the performance of team-taught pupils on these tests be at least equal to that of pupils in self-contained classes—notwithstanding claimed or demonstrated advantages of the team method with respect to pupil attitude, independence, creativity, etc. In the author's opinion this requirement can always be met when it is taken seriously. Then the innovator may in good conscience pursue and elucidate the broader benefits of team teaching that have thus far proved so resistant to convincing demonstration.

REFERENCES

ANDERSON, R. H. Team teaching in the elementary and secondary schools. In A. de Grazia & D. A. Sohn (Eds.), *Revolution in teaching: New theory, technology, and curricula.* New York: Bantam Books, 1964.

BORG, W. R., & BRITE, L. R. Teacher's perceptions of team teaching. *California Journal of Educational Research,* 1967, **18,** 71–81 & 101.

BOYLES, M. P. *A comparative study of pupil achievement and adjustment using instructional teams composed of teachers, teacher assistants, and teacher aides.* (Doctoral dissertation, University of Georgia) Ann Arbor, Mich.: University Microfilms, 1968, No. 68-5034.

BURNINGHAM, G. L. *A study and evaluation of the team teaching of the fourth grade at Woodstock Elementary School.* (Doctoral dissertation, University of Utah) Ann Arbor, Mich.: University Microfilms, 1968, No. 68-11, 905.

FLANDERS, N. A., MORRISON, B. M., & BRODE, L. E. Changes in pupil attitudes during the school year. *Journal of Educational Psychology,* 1968, **50,** 334–338.

GEORGIADES, W., & BJELKE, J. Evaluation of English achievement in a ninth grade, three-period, team-teaching class. *California Journal of Educational Research,* 1966, **17,** 100–112.

KNOX, D. M. *An experimental study of the effect of a team teaching program upon certain selected variables (achievement—anxiety—social relations).* (Doctoral dissertation, Western Reserve University) Ann Arbor, Mich.: University Microfilms, 1966, No. 66-8011.

SHARKAN, W. W. *An evaluation of the team organization plan of staff utilization in relationship to the educational development of students in the* *junior high schools of Allentown, Pennsylvania.* (Doctoral dissertation, Pennsylvania State University) Ann Arbor, Mich.: University Microfilms, 1963, No. 63-3080.

VELDMAN, D. J. *FORTRAN programming for the behavioral sciences.* New York: Holt, Rinehart & Winston, 1967.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

ZWEIBELSON, I., BAHNMULLER, M., & LYMAN, L. Team teaching and flexible grouping in the junior high-school social studies. *The Journal of Experimental Education*, 1965, **34**, 20–32.

(Received January 3, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 2, 117–122

STIMULUS COMPLEXITY AND INTRAMODAL REACTION TIME IN RETARDED READERS[1]†

MARION BLANK,[2] THOMAS J. HIGGINS, AND WAGNER H. BRIDGER
Albert Einstein College of Medicine

The present research was conducted to determine the possible role of stimulus differences in the reported difficulty of cross-modal reaction time in retarded readers. Using an intramodal paradigm involving a "simple" and a "complex" visual stimulus, 20 retarded third-grade readers were found to have significantly longer reaction times than 20 normal third-grade readers. The pattern of response was similar for both reading groups in that the complex stimulus both initially caused more difficulty than the simple stimulus and it led to difficulty in shifting to the subsequent stimulus. The findings suggest that poor cross-modal performance cannot be attributed to deficiencies in cross-modal perception but rather may depend upon the extent of the physical differences that exist between the stimuli, whether within or across modalities.

Cross-modal behavior has begun to be viewed as an area of possible major significance in the learning difficulty of the retarded reader (Beery, 1967; Birch & Belmont, 1964; Katz & Deutsch, 1963; Muehl & Kremenak, 1966; Sterritt & Rudnick, 1966). This term encompasses a wide variety of tasks which involve the presentation of information in two or more sensory modalities (e.g., visual, auditory). While such tasks are in some ways analogous to the reading process, the exact nature of the relationship between reading retardation and cross-modal learning has not been determined. Because of the importance of the concept of perceptual dysfunctioning in reading retardation (Benton, 1962), however, poor cross-modal learning has generally been viewed as a basic deficiency in intersensory perception.

The potential usefulness of this interpretation rests on the assumption that the deficit is specifically limited to intersensory behavior and does not simply reflect generally poor perceptual learning. Unfortunately, as pointed out by Bryant (1968), many of the studies in this area have failed to include the intramodal control conditions necessary for distinguishing between these alternatives. Even when intramodal controls are included, however, their adequacy is subject to question since the stimuli in the intramodal and intermodal conditions are often not equated on essential characteristics. For example, a cross-modal task which has aroused considerable interest requires the child to determine which visual dot pattern (e.g., oo o / oo oo / oooo) corresponds to an auditory sequence (e.g., tap, tap—long pause—tap, tap) (Birch & Belmont, 1964). It should be noted that the auditory stimuli must of necessity be presented temporally. Nevertheless, in the visual-visual intramodal control condition this factor has generally been overlooked and the visual stimuli are presented only on a spatially organized basis (Beery, 1967; Birch & Belmont, 1964; Muehl & Kremenak, 1966). When this factor is controlled by presenting the visual stimuli in a temporal rather than a spatial order (e.g., flashes of lights rather than sets of dots), then the same deficiency that is found in cross-modal learning is presented in intramodal learning as well (Blank & Bridger, 1966; Blank, Weider, & Bridger, 1968).

The failure to equate the intramodal and cross-modal stimuli on essential characteristics seems relevant to another intersensory

---

[1] This research was supported by United States Public Health Service Grant Nos. MH 08498, K3-MH10,749, and 5-KO5-MH04177. The authors express their appreciation to the New York City Board of Education for their cooperation in this research.

[2] Requests for reprints should be sent to Marion Blank, Department of Psychiatry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461.

task that has been used with retarded readers; namely, cross-modal reaction time (Katz & Deutsch, 1963; Raab, Katz, & Deutsch, 1960). In this task it was found that retarded readers took considerably longer than normal readers to respond to a stimulus if they had to shift from one modality to another (cross-modal), but did not differ when having to respond to a difference stimulus within the same modality (intramodal). The intramodal stimuli were a red versus a green light in the visual modality and a high versus a low tone in the auditory modality. The results were interpreted to mean that the retarded reader's difficulty was due to deficiencies in cross-modal transfer and not to stimulus differences per se. This interpretation, however, overlooks the possibility that the intramodal stimuli were much more likely to be closer on a gradient of generalization than was either of the visual stimuli to the auditory stimuli. As a result, the poorer cross-modal performance may have been due not to differences in intersensory ability but rather to the failure of the retarded reader to adapt quickly to markedly different stimuli, regardless of the modality of presentation. The possibility, therefore, exists that if the differences between stimuli within a modality are made extreme, the intramodal performance of the retarded readers would be comparable to their cross-modal performance.

Accordingly, the present research was designed to present an intramodal reaction time task in which stimuli varied markedly across a number of dimensions. Two stimuli were used; one was a unidimensional, white light (termed the "simple" stimulus) and the other was a multidimensional, colored cartoon picture (termed the "complex" stimulus). In an effort to determine whether this intramodal task yielded results comparable to the cross-modal task, a number of factors shown to be significant in the latter situation were assessed. These included the effects of (*a*) practice (*b*) shifting, and (*c*) stimulus differences. In particular, it was predicted that:

1. Retarded readers would display longer reaction times than normal readers on an intramodal situation involving two markedly different stimuli.
2. Retarded readers would display significantly longer reaction times than normal readers to both the simple and complex visual stimuli; however, the complex stimulus would cause the retarded readers relatively more difficulty than the simple stimulus.
3. Both groups will demonstrate significant improvement with practice so that the reaction time in the second half will be significantly more rapid than that in the first half.
4. Comparable to the difficulty found cross-modally, retarded readers will find it harder to shift to a different intramodal stimulus than to continue to respond to the same stimulus.

METHODS

Task and Apparatus

The basic task was that of a reaction time situation in which the subject had to withdraw his finger as quickly as possible upon the presentation of either of two visual stimuli. The total task comprised 150 trials with the first 75 trials being separated from the second 75 by a 1-minute rest period. The stimuli were presented through a reaction time apparatus (Katz & Deutsch, 1963; Sutton, Hakerem, & Zubin, 1961) modified to present two different visual stimuli. The apparatus, 12 × 8 × 7 inches, contained two 3-square-inch openings. Through one of the squares, a plain white light could be presented; through the other, a complex multicolored but predominantly orange cartoon picture (of a rabbit standing in front of a house in the woods) could be shown. When the apparatus was not illuminated, no difference could be discerned between the two openings.

The apparatus was programmed according to a predetermined random schedule which was standardized for all subjects. Each of the two stimuli were presented an equal number of times and each could be displayed from one to six consecutive trials. The time interval between each trial varied randomly between 2 and 5 seconds.

Subjects

The sample, composed of 20 retarded and 20 normal third-grade readers, was selected from a New York City elementary school attended mainly by children from the lower and lower-middle socioeconomic class. Third-grade children were used since in previous research (Katz & Deutsch, 1963) no difference in intramodal performance was found at this age between normal and poor readers. The subjects, 11 boys and 9 girls in the normal group and 9 boys and 11 girls in the retarded group, had a mean age of 8 years, 11 months. All testing was conducted near the end of the school year when normative performance is about at fourth-grade level. Through their Metropolitan Achievement Test performance, subjects were classified as retarded readers if they tested at a year or more below grade level and as normal readers if they performed at least at grade level. The mean reading level of the retarded readers was 2.0 and of the normal readers was 4.7.

Procedure

Each subject was individually tested in a room in his school. He was seated in front of the apparatus, shown the two stimuli, and instructed to raise his index finger "as quickly as possible" whenever any of the two stimuli were presented. Following the verbal instructions, the experimenter gave a demonstration of the procedure to each subject. The stimulus was terminated by the subject's response and the reaction time, in milliseconds, was measured from the stimulus onset to the lifting of the finger.

RESULTS

Table 1 presents the group results for both the retarded and normal readers according to both stimulus presented and first versus second half of performance. Two three-way analyses of variance were performed on the data. The first was designed to assess the variables of reading level (normal, retarded), type of stimulus (simple, complex), and practice effects (first versus second half of performance). The second analysis was designed to assess the effect of intramodal shifting. Because the distributions were skewed and the variance heterogeneous, a square root transformation for each score was performed before applying parametric statistics to these data.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF REACTION TIMES IN MILLISECONDS TO INTRAMODAL TASK

Group	Visual stimulus			
	Simple		Complex	
	First half	Second half	First half	Second half
Normal readers				
M	664.6	643.8	689.6	625.0
SD	156.8	98.2	191.5	94.8
Retarded readers				
M	813.9	764.6	845.9	773.6
SD	239.1	234.7	286.2	244.8

TABLE 2

RESULTS OF THREE-WAY ANALYSIS OF VARIANCE INVOLVING READING LEVEL, STIMULUS COMPLEXITY, AND PRACTICE EFFECTS

Source	*df*	*MS*	*F*
Reading level (A)	1	249.80	5.82*
Error	38	42.86	
First versus second half (B)	1	28.34	8.26**
A × B	1	.57	<1
Error	38	3.43	
Stimulus complexity (C)	1	.57	1.00
A × C	1	.33	<1
Error	38	.57	
B × C	1	3.29	7.31*
A × B × C	1	.50	1.11
Error	38	.45	

* $p < .05$.
** $p < .01$.

The first three-way analysis, involving repeated measurements on two variables, supports Hypotheses 1 and 3 in that the main effects of reading level and practice were significant (see Table 2). Thus, normal readers consistently responded faster than retarded readers on this intramodal task and both groups showed marked improvement from the first to the second half. While stimulus complexity did not yield a significant main effect, it did interact with practice, in that the complex stimulus caused significantly longer reaction time than the simple stimulus in the first half ($t = 2.57$, $df = 38$, $p < .05$), but not in the second half ($t = 1.14$, $df = 38$, $p > .20$). This indicates that practice resulted in the elimination of those differences in reaction time which were caused by the difference between the stimuli. These results further suggest that relatively rapid reaction time to the complex stimulus was more dependent on practice than was the reaction time to the simple stimulus. Contrary to predication, no significant interactions occurred between reading level and stimulus complexity (i.e., the complex stimulus did not cause relatively greater difficulty for the retarded reader than did the simple stimulus). However, it should be noted that the results tend in the predicted direction in that the normal readers so decreased their latencies that by the second half, their mean response was shorter to the complex than to the simple stimulus. It is possible that a significant effect was not obtained because the complex stimulus was not as sufficiently differentiated from the simple stimulus as it might have been (e.g., different and varied patterns might have been presented in a random order in the locus where the single complex stimulus appeared).

The results were also analyzed to determine the effects on reaction time of having to shift to a different stimulus versus continuing to respond to the same stimulus (see Table 3). This was done to get a better understanding of the process of shifting since cross-modal work indicated that the trials involving a shift across modalities best discriminated the reading groups. The analogous measure in the present study was having to shift to a different stimulus within the modality. A second three-way analysis of variance was conducted since the data for this analysis had to be totally reorganized and new means computed for each subject. As shown in Table 4, both analyses yielded almost identical results in that reading level again was highly significant and stimulus complexity did not approach significance. In addition, contrary to the hypothesis, reading level did not interact with either of the other two factors. Thus, as in the first analysis of variance, while retarded readers had slower overall reaction times, the pattern of their responses was very similar to that of the normal readers. Although the main effect of shifting (responding to the same versus a different stimulus) was not significant, there was a significant interaction between this factor and stimulus complexity. Thus, when the children were required to continue to respond to the same stimulus, their reaction time to the simple stimulus was significantly more rapid than to the complex stimulus ($t = 2.61$, $df = 38$, $p < .05$). In addition, the reaction times to both the simple and complex stimulus were more rapid when the preceding stimulus was simple rather than complex (i.e., for the comparison between simple preceded by simple versus simple preceded by complex, $t = 2.15$, $df = 38$, $p < .05$ and for the comparison of the complex preceded by the simple versus complex preceded by complex, $t = 2.07$, $df = 38$, $p < .05$). These results suggest that the physical properties of the stimuli affected the ease with which attention could be mobilized for the succeeding stimulus. In the present study this was reflected in the difficulty of shifting from the complex stimulus, regardless of what the succeeding stimulus was.

TABLE 3

Mean Reaction Time to Same Stimulus versus Shifting to Different Stimulus

Group	Preceding stimulus	Reaction time	
		Simple	Complex
Normal readers	Simple	658.4	644.7
	Complex	653.0	664.5
Retarded readers	Simple	780.0	788.8
	Complex	802.7	798.8

TABLE 4

Results of Three-Way Analysis of Variance Involving Reading Level, Stimulus Complexity, and Shifting

Source	*df*	*MS*	*F*
Reading level (A)	1	246.51	6.29*
Error	38	39.13	
Stimulus Complexity (B)	1	.16	<1
A × B	1	.07	<1
Error	38	.68	
Shifting Same-Different (C)	1	.00	<1
A × C	1	.29	<1
Error	38	.50	
B × C	1	2.99	8.79**
A × B × C	1	.01	<1
Error	38	.34	

* $p < .05$.
** $p < .01$.

Discussion

The results indicate that in contrast to previous reports, (Katz & Deutsch, 1963) retarded readers display longer intramodal reaction times than do normal readers if the stimuli within the modality are markedly different. These findings should not be interpreted to mean that there would have been no differences between cross-modal and intramodal learning, had the cross-modal situation been available. The findings do imply, however, that one cannot attribute poor cross-modal performance simply to deficiencies in cross-modal perception since the results may well depend upon the extent of the physical differences that exist between the stimuli, whether within or across modalities. It is recognized that there may have been, and likely were, IQ differences between the normal and retarded readers. However, such differences also existed in the subjects of the Katz and Deutsch study and poorer intramodal performance was not found. It might also be argued that the introduction of a complex stimulus resulted in the retarded readers having difficulty and that these results would not have occurred had the present study been confined to simple stimuli such as the red and green lights originally used in this research. It should be noted, however, that the red and green lights were specifically introduced not for their simplicity, but rather to control for the problem of stimulus differences per se. In addition, if restrictions are imposed as to the type of stimulus that can be presented within this context, then it becomes impossible to study the role played by stimulus differences in cross-modal shifting.

The problem, however, remains of determining the reasons for the poor performance of the retarded readers on this intramodal task. In the cross-modal research the explanation for this phenomenon has generally involved the concept of attention—retarded readers are deemed to have difficulty in shifting attention in accord with the demands of the situation. Data from the present study concerning the differential difficulty in shifting from the complex as opposed to the simple stimulus offers support for the importance of attentional mechanisms. However, it should not be assumed, as it easily might be, that the poor performance of the retarded readers represents a failure in attention. The difficulty the retarded readers had in dealing with the complex stimulus suggests that the children were affected by the physical characteristics of the stimuli. This result raises an interesting possibility, namely, that the retarded readers had difficulty in shifting not because of a failure in attention, but rather because of too great attention to the stimulus characteristics. As Sutton et al. (1961) noted in their design of this task, stimulus information must be ignored if the shifting is to be effective. In most work to date with this paradigm, it has been assumed that the simplicity of the task would ensure that this would occur: "Since the task is a simple reaction, requiring no discrimination in the perception of the signal nor in the execution of the response, the ipsimodal and crossmodal reaction times should not differ, unless the shift in modality *per se*—the phenomenon *E* was investigating—introduces a difference. [Sutton et al., 1961, p. 225]." While it is possible for the task to be carried out without having to discriminate among the stimulus characteristics, in fact this may be a difficult thing to do since the information to be ignored is exactly the kind of information that is most arousing and salient (i.e., the changing lights). In addition, it is not simply that stimuli must be ignored. Rather the basic demand is one of "selective inattention"; that is, at one and the same time the child is asked to "ignore all qualities of the stimuli you see" but "pay attention to the fact that a stimulus is present." In effect, this demands a complex set from the child and the poor performance of the retarded reader may be the result of his difficulty in adopting this set. A similar failure in adopting the appropriate set to a seemingly simple perceptual task was found in previous research (Blank, 1968).

This explanation might be relevant in accounting for the improvement with practice found in both this and the Katz and Deutsch (1963) study; that is, as the child becomes habituated to the stimuli, he is less drawn to their physical properties. As a result, he can begin to ignore their characteristics and concentrate on the hoped for aim of the task, namely a rapid reaction to the onset of the stimulus. While support for this hypothesis from the present study is limited, this approach permits one to derive testable hypotheses which may help decide upon alternative explanations for the poorer functioning of the retarded reader. For example, if the retarded reader is overattentive rather than underattentive to the properties of physical stimuli, then habituating him to the stimuli (e.g., by simply displaying the sequences of stimuli without allowing the children to practice the motor response) should lead to a dimunition of the differences between the retarded and normal reader. Conversely, constant changes in the stimuli to be presented should lead to an increase in the differences between the retarded and normal readers.

Resolution of these issues is not only of intellectual interest, but may also be of practical importance. The concept of "subtle impairments of perceptual functioning" (Katz & Deutsch, 1963, p. 327) has gained increasing acceptance as a major factor in reading retardation. As the authors' previous work has shown, however, the handling of seemingly simple perceptual tasks often requires higher level cognitive control (Blank, 1968; Blank & Bridger, 1966, 1967; Blank, Weider, & Bridger, 1968). Therefore, it is important to determine through further research whether the retarded readers' difficulty on the reaction time paradigm is caused by an inability to shift attention, within or across modalities, or by a failure to adopt the set appropriate to this situation (i.e., ignore the "attraction" of the perceptual characteristics). If this phenomenon is important to the reading process, rather than simply a correlate of reading retardation, it is evident that effective remediation will depend upon which factors are in fact bound to be responsible for the deficit.

REFERENCES

BEERY, J. Matching of auditory and visual stimuli by average and retarded readers. *Child Development*, 1967, **38**, 828–833.

BENTON, A. Dyslexia in relation to form perception and directional sense. In J. Money (Ed.), *Reading disability*. Baltimore: John Hopkins Press, 1962.

BIRCH, H., & BELMONT, L. Auditory-visual integration in normal and retarded readers. *American Journal of Orthopsychiatry*, 1964, **34**, 852–861.

BLANK, M. Cognitive processes in auditory discrimination in normal and retarded readers. *Child Development*, 1968, **39**, 1091–1101.

BLANK, M., & BRIDGER, W. H. Deficiencies in verbal labeling in retarded readers. *American Journal of Orthopsychiatry*, 1966, **36**, 840–847.

BLANK, M., & BRIDGER, W. H. Perceptual abilities and conceptual deficiencies in retarded readers. In J. Zubin & G. Jervis (Eds.), *Psychopathology of mental development*. New York: Grune & Stratton, 1967.

BLANK, M., WEIDER, S., & BRIDGER, W. H. Verbal deficiencies in abstract thinking in early reading retardation. *American Journal of Orthopsychiatry*, 1968, **38**, 823–834.

BRYANT, P. E. Comments on the design of developmental studies of cross-modal matching and cross-modal transfer. *Cortex*, 1968, **4**, 127–137.

KATZ, P. A., & DEUTSCH, M. Relation of auditory-visual shifting to reading achievement. *Perceptual and Motor Skills*, 1963, **17**, 327–332.

MEUHL, S., & KREMENAK, S. Ability to match information within and between auditory and visual sense modalities and subsequent reading achievement. *Journal of Educational Psychology*, 1966, **57**, 230–239.

RAAB, S., KATZ, P., & DEUTSCH, M. Perceptual shifting and set in normal school children of different reading achievement levels. *Perceptual and Motor Skills*, 1960, **10**, 187–192.

STERRITT, G., & RUDNICK, M. Auditory and visual perception in relation to reading ability in fourth grade boys. *Perceptual and Motor Skills*, 1966, **22**, 859–864.

SUTTON, S., HAKEREM, G., & ZUBIN, J. The effect of shift of sensory modality on serial reaction-time: A comparison of schizophrenics and normals. *American Journal of Psychology*, 1961, **74**, 224–232.

(Received January 8, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 2, 123–131

INDIVIDUAL DIFFERENCES IN VISUAL AND AUDITORY MEMORY[1]†

ARTHUR R. JENSEN[2]
University of California, Berkeley

Two groups were given visual and auditory forms of a digit memory test in a counterbalanced order (auditory-visual group and visual-auditory group) under conditions of immediate and 10-second delayed recall. Two control groups were given exclusively either the visual or auditory test. Auditory memory was better than visual for immediate recall; the reverse was true for delayed recall, the interaction being significant beyond $p < .001$. The correlations between individual differences in auditory and visual memory, after correction for attenuation based on reliabilities obtained from the control groups, did not significantly differ from unity for either immediate or delayed recall. Thus there was no evidence for individual differences as a function of sensory modality of the input. On the other hand, there was a significant ($p < .001$) interaction between subjects and time of recall (immediate versus delayed).

Are there visual learners and auditory learners? Do some individuals learn and remember better in the one sensory modality than in the other? It is at least part of the general folklore of psychology that such auditory and visual types exist. Wechsler (1958), for example, in discussing the digit memory span test, states: "It is certainly true that a good auditory memory does not go with a good visual memory [p. 131]."

Interestingly enough, despite the fact that there is a venerable literature pertaining to this question, an effort to find any real empirical support in psychological research for Wechsler's statement has come to naught. Though the issue has been often discussed and even researched, it has apparently never been subjected to a proper experiment—one so designed as to be capable of answering the key question, at least for one type of memory. We know, of course, that no single experiment could answer the question for all types of learning and memory, since there are large task-specific sources of variance which interact with individual differences. Therefore, the present paper is concerned with this question only with respect to short-term memory, specifically memory for digits, a traditional part of standard psychometric tests such as the Stanford-Binet and the Wechsler intelligence scales.

The problem really comes down to two questions: (*a*) Do people in general learn or remember more effectively when the material is presented visually or aurally? (*b*) Do some individuals favor one sensory modality over another in learning and remembering?

Most studies have tried to answer the first question; few have tried to answer the second. The results are far from unanimous. Some investigators have found auditory memory superior to visual (Binet, 1894; Koch, 1930; Münsterberg & Bingham, 1894). Others have found visual superior to auditory (Hawkins, 1897; Henmon, 1912; Kirkpatrick, 1894; O'Brien, 1921; Worchester, 1925). Simultaneous presentation of the material visually and aurally always produced the best recall (Binet, 1894; Koch, 1930; Münsterberg & Bingham, 1894). Day and Beach (1950), who reviewed the entire literature up to 1950, concluded that, in terms of a box score, visual memory is generally better than auditory. But a box score of experimental outcomes is, of course, scientifically a confession of ignorance. It means the problem is complex and the conditions that produce one or another outcome must be precisely specified. Interactions of

---

[1] The research reported in this paper was performed under a contract (Project No. 1867) with the Cooperative Research Program of the Office of Education, United States Department of Health, Education, and Welfare.

[2] Requests for reprints should be sent to Arthur R. Jensen, School of Education, University of California, Berkeley, California 94720.

TABLE 1

STUDIES OF THE CORRELATION BETWEEN VISUAL AND AUDITORY MEMORY

Investigator	Material	N	Reliability		Correlation r_{AV}
			r_{AA}	r_{VV}	
Gates (1916c)	Digits	172			.57
Gates (1916b)	Digits	165			.62
Hao (1924)	Digits	83	.52	.83	.39
Wissler (1901)	Digits (written)	144			.29
	Digits (placed)	144			.39
Kitson (1917)	Various	40			.09–.54
Gates (1916a)	Various	197			.07–.41
Carey (1914)	Various (pooled)	150	.68	.64	.44

Note.—The author is indebted to John J. Geyer for compiling the information in this table.

the key variable (sensory modality) with other experimental variables could account for seemingly contradictory results in different studies, as could the interaction of the key variable with subject variables or individual differences. For example, although the majority of studies show that adults do better on visual than on auditory memory tasks, it appears that the reverse is true for children (Abbot, 1909; Hawkins, 1897; MacDougall, 1904). And visual short-term memory has been found to decline more rapidly with age (over 40) than auditory memory (McGhie, Chapman, & Lawson, 1965). In brief, the question of the superiority of visual or auditory memory cannot be given a general answer that covers all conditions and types of subjects. An experimental analysis of the specific conditions that make for the superiority of one or the other is needed.

What about the second question, that of individual differences? It should be answerable for a specified type of memory task. Kay (1958) used a paired-associate test consisting of familiar words and found that 45% of the subjects performed about equally in both modes of presentation, 12% did significantly better in the visual, and 4% did better in the auditory. If the same kind of material is presented to a group of subjects in each of two modalities independently, all aspects of the presentation being held constant except sensory modality, then the correlation between subjects' performance in the two modalities should be significantly less than 1.00 (after correction for attenuation), if the hypothesis of individual differences in the relative efficacy of the visual and auditory modalities in short-term memory is to be upheld. Several studies have been conducted in this vein. They are summarized in Table 1. The raw correlations (r_{AV}) between auditory and visual performance, being generally quite low, would give the impression that individual differences in auditory and visual memory are not highly correlated. But in most cases, we do not know the reliabilities (r_{AA} and r_{VV}) of the auditory and visual memory tests, and therefore cannot interpret the magnitude of the correlation r_{AV}. In the two instances where the reliabilities (r_{AA} and r_{VV}) are given, we do not know if they were obtained by a method that parallels the method for obtaining r_{AV}. Split-half reliability, for example, is different from test–retest reliability in theory and usually in fact, and so it would make little sense to correct r_{AV} for attenuation by using either type of reliability coefficient arbitrarily. The lack of any real answer to our question revealed by these data thus highlights the minimum requirements of any experiment addressed to the problem.

Essentials of Experimental Design

A proper test of the hypothesis must be designed to answer the question of whether the correlation (corrected for attentuation) between auditory and visual memory is significantly less than 1.00, that is,

$$\frac{r_{AV}}{\sqrt{r_{AA}r_{VV}}} < 1.00$$

This implies: (*a*) Reliability estimates (r_{AA} and r_{VV}) that are at least as dependable as the correlation between auditory and visual performance (r_{AV}) must be obtained. (*b*) Reliability coefficients must be in every way parallel to the r_{AV} correlation, that is, both the total amount of practice and the intervening time interval in the subjects' performance for obtaining the auditory-auditory and the visual-visual correlations (r_{AA} and r_{VV}) must be identical to the procedures for obtaining the auditory-visual correlation (r_{AV}). This can be achieved only by using independent groups of subjects randomly sampled from the same pool for determining the reliabilities of the visual and auditory tests, and still another group for determining the correlation between the visual and auditory tests. (*c*) The possibility of practice in one modality transferring differentially to the other must be counterbalanced by using two groups to obtain the correlation between modalities, one which practices auditory first and then visual, and one which does the reverse. The two correlations that result are designated r_{AV} and r_{VA}, respectively. (*d*) The testing procedure should be identical in all respects except for the sensory modality of the presentation.

The present experiment was designed to fulfill these requirements. In addition to sensory modality, one other experimental variable was introduced: immediate recall versus a 10-second delay in recall. Since time of recall cannot be precisely controlled in individual subjects, it is preferable to attempt to assess the effect of the variable by manipulating it and determining if it interacts with the subjects and with sensory modality.

METHOD

Subjects

The subjects were 150 undergraduate university students, 19 to 23 years of age, in an introductory educational psychology course. They were randomly assigned to four groups.

Group AA ($n = 50$). Subjects were given equivalent forms of the auditory digit memory test on each of 2 days.

Group VV ($n = 50$). Subjects were given equivalent forms of the visual digit memory test on each of 2 days.

Group AV ($n = 25$). Subjects were given the auditory digit memory test on the first day (Day 1) and the visual test on the second day (Day 2).

Group VA ($n = 25$). Subjects were given the visual test on Day 1 and the auditory test on Day 2.

Procedure

Apparatus. The subjects were tested alone in a sound-protected room which was barren except for the simple apparatus essential for the testing. The experimenter could observe the subject from an adjoining room through a one-way window. On the visual test, digits (2 inches high) were presented on an in-line display unit placed approximately 2 feet from the seated subject, at eye level. Rate of digit presentation, stimulus selection, and all time intervals were preprogrammed on a punched tape and were automatically controlled by an apparatus described in detail elsewhere (Jensen, Collins, & Vreeland, 1962). For the auditory test, the in-line display unit was replaced by a speaker which presented the digits (and other signals involved in the procedure) in a normal, clear female voice. The tape recording was made to precisely the same pacing rate and time intervals as the visual presentation by having the speaker read aloud the entire automatically presented visual test into a tape recorder. In front of the subject, who was seated at a table, there was a response console tilted at a 30° angle. Directly in the center of it was a clipboard holding the specially prepared forms on which the subject would write his response after each series of digits had been presented. Three inches away from each side of the clipboard were two pushbuttons (¾-inch in diameter) protruding from the console; the right-hand button was labeled with a red plus (+) sign and the left-hand button was labeled with a red minus (−) sign. To prevent the subjects' writing during the presentation of the digit series and during the delay interval prior to recall, the subject was required to keep his index fingers poised on the pushbuttons. (Failures to do so at the prescribed times were automatically recorded by the apparatus.)

Materials

The digit series, comprised of numerals 1 through 9 (zero was never used), were originally produced by a computer; the sequence was always random, with two exceptions: (*a*) no digit was ever repeated in the series, and (*b*) no two digits ever occurred in the normal numerical order in the forward direction, such as 3–4 or 7–8.

There were two conditions of recall, immediate and delayed; in the delayed condition the subject received the signal to write his response 10 seconds after the last digit of the series had been presented. The delay interval was filled by a random sequence of red plus and minus signs, at a 1-second rate, on the screen (in the visual presentation) or spoken (in the audio presentation); the subject was required to press the corresponding plus and minus pushbuttons as these symbols appeared on the screen or issued from the speaker. The subject's errors and omissions of pushbutton responses were automatically tabulated on counters in the control unit, so it was possible to tell if the subjects were obeying instructions.)

Each digit series was preceded by the sound of a "bong" as a signal to pay attention, and after the digit presentation a "bong" signaled for the subject to write his response. (The same "bong" was common to both the visual and auditory presentations.)

Instructions

The task was explained to the subject with the aid of a wall chart which showed the two conditions (immediate and delayed) and the design of the whole procedure, as follows:

Immediate Recall

Events	*Time*
bong (ready signal)	1 second
blank	1 second
digits (2 to 9)	2–9 seconds
bong (signal to write)	1 second
blank (for writing response)	13 seconds
bong (ready signal)	1 second
etc.	

Delayed Recall

Events	*Time*
bong (ready signal)	1 second
blank	1 second
digits (2 to 9)	2–9 seconds
blank	1 second
+ and − (random order)	8 seconds
bong (signal to write)	1 second
blank (for writing response)	13 seconds
bong (ready signal)	1 second
etc.	

There were five replications of the test in each day's session. Thus there were 8 (series lengths) × 2 (immediate or delayed recall) × 5 (replications) = 80 digit series presented in each test.

To insure that all subjects understood the instructions and procedure, the test proper was preceded by a practice test consisting of eight series (four of each condition, immediate and delayed); the digit series was never larger than five in the practice test.

The equivalent forms (1 and 2) of the test were identical in all respects except that different digit series were used in each. The visual and auditory tests also were identical (for both Forms 1 and 2) except in mode of stimulus presentation.

The subjects were tested on Forms 1 and 2 on each of 2 days, Day 1 and Day 2, 1 week apart, at the same hour.

Measures

The subjects were instructed to write down in correct sequence as many of the digits as they could recall. The answer sheet contained 80 series of 9 boxes for the subject to write his responses. The subject's score is the total number of digits recalled in the correct positions.

The possible total number of correct responses for 1 day's test session (of approximately 50 minutes) was thus the sum of series lengths $2 + 3 + \cdots + 9 = 44 \times 2$ (delayed or immediate) $\times 5$ (replications) $= 440$. The subjects' answer sheets were directly punched on IBM cards and scoring was done by computer.

RESULTS

Effects of Sensory Modality

Table 2 presents the means and standard deviations of digit recall scores for the four groups in the study. All mean differences between the main effects of auditory and visual, between immediate and delayed recall, and between Day 1 and Day 2 (i.e., the general practice effect) are revealed by analysis of variance to be significant beyond the .001 level. Day 2 performance is significantly superior to Day 1, indicating a practice effect. Immediate recall is nearly always superior to delayed recall; only 2 of the 150 subjects showed either no overall difference or a difference in favor of delayed recall. Overall, visual was superior to auditory, but this fact must be viewed in relation to

TABLE 2

SUMMARY STATISTICS ON THE FOUR GROUPS IN THE EXPERIMENT

Measure	Visual (Group VV)				Auditory (Group AA)				Visual-auditory (Group VA)				Auditory-visual (Group AV)			
	Day 1		Day 2		Day 1		Day 2		Day 1		Day 2		Day 1		Day 2	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Immediate (I)	172.22	22.25	182.94	20.69	180.37	20.05	184.39	20.02	182.84	21.60	181.88	22.33	169.28	23.46	177.96	26.38
Delayed (D)	152.66	30.16	171.26	25.41	136.00	25.68	146.29	28.32	163.00	29.61	137.60	31.48	129.48	27.84	156.60	34.59
Difference (I − D)	19.56	17.32	11.68	17.64	44.37	19.44	38.10	21.48	19.84	16.62	44.28	13.60	39.80	19.02	21.36	18.64
Relative difference[a]	11.38	10.78	6.10	9.29	24.20	10.52	20.31	11.44	11.00	9.90	24.72	9.91	23.24	11.81	12.40	12.55

Note.—$n = 50$ for Group VV and for Group AA, $n = 25$ for Group VA and for Group AV.

[a] The formula for the relative difference is $[(I - D)/I] \times 100$.

the marked interaction between sensory modality and time of recall.

The interaction between sensory modality (auditory versus visual) and time of recall (immediate versus 10-second delay) is significant beyond the .001 level. The form of this interaction is shown in Figure 1. For immediate recall, auditory was superior to visual ($p < .05$) and for delayed recall, visual was markedly superior to auditory ($p < .001$).

Transfer or practice effects from visual to auditory and from auditory to visual practice unfortunately cannot be properly assessed from these data. Differences in difficulty level of the auditory and visual conditions on Day 1 are confounded with Day 2 performance, such that comparisons of Day 2 scores between Groups VA and AA and between Groups AV and VV cannot be rigorously interpreted with respect to the relative amounts of transfer from one sensory modality to the other. This transfer or practice effect could be properly assessed only if the auditory and visual tasks were somehow experimentally equated in difficulty. (Statistical equating, as by analysis of covariance, would be an incorrect procedure in this case.)

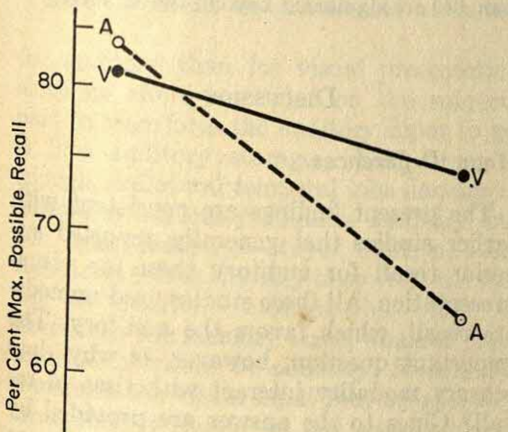


FIG. 1. Recall (as percentage of maximum possible recall) of auditory (A) and visual (V) digit series, under two conditions of recall: (*a*) immediately following presentation of the digit series, and (*b*) following a 10-second delay after the presentation of the series. (Series varied in length from two to nine digits presented at a rate of 1 digit per second.)

Serial Position Effects

Since serial position effects (i.e., probability of correct recall of a digit as a function of its position in the series) have some bearing on the theoretical interpretation (see Discussion section) of the interaction shown in Figure 1, the serial position data are given in Tables 3 and 4 as probability of correct recall of each digit at each position in the series.

Individual Differences

The principal question this study was designed to answer is, Do some persons have better short-term memory in one modality than in another? If the true-score correlation between modalities is significantly less than unity, the answer to this question is yes. If the true-score correlation does not differ significantly from unity, it cannot be concluded that individual differences in recall interact with sensory modality of the presentation.

The appropriate reliability coefficients for the visual and auditory tasks are obtained from the correlations (r_{AA} and r_{VV}) between Day 1 and Day 2 for Group AA and Group VV, for immediate recall and delayed recall separately. These reliability coefficients are then used to correct the correlations (r_{AV} and r_{VA}) between auditory and visual tasks for attenuation to obtain the true-score correlation between modalities. The true-score (i.e., corrected) correlations (r') are $r'_{AV} = r_{AV}/\sqrt{r_{AA}r_{VV}}$ and $r'_{VA} = r_{VA}/\sqrt{r_{AA}r_{VV}}$. The results are shown in Table 5. Since none of the corrected correlations is significantly less than unity, these data do not support the hypothesis that individual differences in memory for digits are a function of the sensory mode (visual versus auditory) of presentation. This is true for both immediate and delayed recall. Even though the mean difficulty level of auditory and visual tasks is reversed in immediate and delayed recall, the rank order of individuals' recall scores remains the same in both modalities. Another way of stating this, in terms of the

TABLE 3

PROBABILITY OF CORRECT RECALL IN VISUAL DIGIT SERIES AS A FUNCTION OF LENGTH OF SERIES, SERIAL POSITION OF DIGIT, AND TIME OF RECALL (IMMEDIATE VERSUS DELAYED)

Series length	Time of recall	Serial position 1	2	3	4	5	6	7	8	9	*M*
2	Immediate	982	987								985
	Delayed	975	962								968
3	Immediate	983	981	985							983
	Delayed	981	965	951							966
4	Immediate	982	979	977	975						978
	Delayed	954	937	919	833						911
5	Immediate	976	956	940	949	961					956
	Delayed	932	850	830	794	826					846
6	Immediate	982	936	907	890	906	921				924
	Delayed	971	899	843	800	772	794				847
7	Immediate	942	907	861	795	779	749	786			831
	Delayed	916	775	691	644	627	558	519			676
8	Immediate	896	778	724	662	594	599	550	557		670
	Delayed	872	764	683	604	536	499	428	444		604
9	Immediate	914	728	623	605	552	505	494	418	458	589
	Delayed	856	733	597	523	464	396	359	311	340	508

Note.—Decimals are omitted. Differences greater than .040 are significant beyond the .01 level.

analysis of variance, is that there is no significant Subjects × Modality interaction.

Interaction of Subjects × Recall Condition

Time of recall (immediate versus delayed), on the other hand, interacts significantly with subjects, both for visual (Group VV: $F = 4.09$, $df = 49/196$, $p < .001$) and for auditory (Group AA: $F = 5.62$, $df = 49/196$, $p < .001$) presentation. Individual differences in immediate and delayed recall have between 60% and 70% of their true-score variance in common. While virtually all subjects show poorer delayed than immediate recall, some subjects show a considerably greater memory decrement after delayed recall than do others. Thus, subjects do not maintain the same rank order of ability in immediate and delayed recall, while they do maintain the same rank order in the aural and visual tests.

DISCUSSION

Mean Differences

The present findings are consistent with earlier studies that generally reported superior recall for auditory than for visual presentation. All these studies used immediate recall, which favors the auditory. The important question, however, is why does sensory modality interact with time of recall? Clues to the answer are provided by recent theory and research on short-term memory.

Sperling (1963) has hypothesized that a primary function of verbal rehearsal in visual memory tasks is to convert information from visual to auditory storage. That is to say, the memory traces are transformed and encoded in an auditory form, regardless of the sensory channel of reception. Rehearsal, therefore, should be less necessary

TABLE 4

PROBABILITY OF CORRECT RECALL IN AUDITORY DIGIT SERIES AS A FUNCTION OF LENGTH OF SERIES, SERIAL POSITION OF DIGIT, AND TIME OF RECALL (IMMEDIATE VERSUS DELAYED)

Series length	Time of recall	Serial position									*M*
		1	2	3	4	5	6	7	8	9	
2	Immediate	997	997								997
	Delayed	984	973								979
3	Immediate	998	997	998							998
	Delayed	981	954	957							964
4	Immediate	996	982	918	929						956
	Delayed	976	926	878	818						900
5	Immediate	997	978	962	969	983					978
	Delayed	917	796	744	672	689					764
6	Immediate	986	932	876	844	896	950				914
	Delayed	940	805	698	628	568	636				712
7	Immediate	980	922	870	805	761	750	858			849
	Delayed	893	700	594	466	428	340	373			542
8	Immediate	921	794	708	639	592	588	615	700		695
	Delayed	817	631	517	459	376	320	255	272		456
9	Immediate	900	733	588	589	469	465	473	508	670	599
	Delayed	830	710	509	409	361	284	247	227	257	426

Note.—Decimals are omitted. Differences larger than .040 are significant beyond the .01 level.

for auditory than for visual presentation, since no effort is needed on the subject's part to transform the auditory input to get it into auditory storage. (Yet, strangely enough, unilateral temporal lobe damage or lobectomy seriously impairs auditory but not visual memory—Meyer, 1961, p. 545.) Rehearsal serves not only to convert the visual to auditory storage, but also to strengthen the memory trace, whether presentation is visual or auditory. If subjects transform the visual input into an auditory form, then it should be less susceptible to visual retroactive interference (the series of red pluses and minuses projected on the screen) than to auditory forms of retroactive interference (a spoken series of "plus" and "minus"). The subjects often report assuming a more passive attitude in the auditory than in the visual presentation; they experience an "echo chamber" effect for immediate recall of auditory digits, and simply read back the "echo" in writing down their answer. But the echo fades rapidly and also is highly susceptible to interference by any other intervening auditory input, and thus there is a greater decrement in the delayed recall of auditory as compared with visual digits.

This interpretation is consistent with Gates' (1916b) finding that subjects tested with lists that exceeded their immediate memory span remembered 25.9% less than

TABLE 5

RAW CORRELATIONS BETWEEN DAY 1 AND DAY 2 FOR THE FOUR GROUPS AND CORRELATIONS CORRECTED FOR ATTENUATION FOR THE CHANGED-MODALITY GROUPS (AUDITORY-VISUAL AND VISUAL-AUDITORY)

Recall condition	Raw correlations				Corrected correlations	
	r_{AA}	r_{VV}	r_{AV}	r_{VA}	r'_{AV}	r'_{VA}
Immediate	.82	.70	.86	.78	1.13	1.03
Delayed	.80	.79	.79	.79	.99	1.00

their span on visual and 36.6% less on auditory tasks. If we assume that the amount recalled is less than the subject's span when the series presented exceeds the span because the subsequent items in the series retroactively inhibit the earlier items, it appears from Gates' results that auditory presentation is more retroactively inhibiting than visual presentation. In the present experiment it is interesting that the digit position with the lowest probability of correct recall (in series of three to nine digits) is, on the average, 2.14 positions from the end of the series for visual, and 2.71 for auditory (see Tables 4 and 5). This is consistent with Gates' (1916b) finding.

Waugh (1960) has suggested that subjects rehearse a digit series (or any other kind of serial list) "cumulatively"—that is, during presentation, subjects repeat to themselves the items from the beginning of the series up to the item being presented. The earlier items are therefore more thoroughly rehearsed, and this would account for the pronounced primacy effect (i.e., superior recall of the first items as compared with the last) typically found in short-term serial memory. Corballis (1966) tested this hypothesis for immediate recall of digits separately under visual and auditory presentation. The experiment also included two conditions of presentation—one favored cumulative rehearsal, the other did not. It turned out there was cumulative rehearsal for visual digits but little, if any, for auditory. Corballis concluded that rehearsal, particularly cumulative rehearsal, is more restricted by auditory than by visual presentation.

The Waugh and Corballis hypotheses should predict a greater primacy effect for the visual than for the auditory digit series, due to greater cumulative rehearsal from the beginning of the list for the visual than for the auditory series. And we should expect this to show up even more for delayed than for immediate recall, since rehearsal should lessen decay of the memory trace. This prediction is not entirely borne out by the present data. An index of the primacy effect was derived for every length of list; it consisted simply of the ratio of first-position correct responses to last-position correct responses. In immediate recall, the mean index was 1.24 and 1.12 for visual and auditory tasks, respectively; but delayed recall showed the opposite: 1.47 for visual versus 1.83 for auditory. Since delay of recall increases the primacy effect for both visual and auditory tasks, it is puzzling, and also contrary to the Waugh-Corballis hypothesis, that the strongest primacy is found for auditory delayed recall.

Individual Differences

The fact that there are no significant individual differences as a function of sensory modality would seem consistent with the hypothesis that the stimuli, regardless of sensory channel, are encoded in a single auditory short-term memory system. Under this condition, and assuming the subjects had no primary visual or auditory defects, the only source of individual differences in visual and auditory memory would be in the effectiveness with which subjects transformed visual input into an auditory memory trace. This can be thought of as a kind of mediation process. It might well be that young children, the mentally retarded, the senile, and certain types of brain-damaged subjects would show marked individual differences in auditory and visual memory span. A college population, on the other hand, by its very nature has been thoroughly, though indirectly, screened for these conditions. Also, one would expect a more or less uniformly high level of development of the rudimentary transformational or mediational skills involved in digit memory among a young, intellectually superior segment of the population. (The present subjects were at least in the upper 10% of all high school graduates.) The present results, therefore, cannot safely be generalized to a population that includes all ages, all levels of intelligence, or organic brain abnormalities.

REFERENCES

ABBOT, E. E. Memory consciousness in orthography. *Psychological Monographs,* 1909, **11**, 127-158.

BINET, A. *Introduction à la psychologie experimentale.* Paris: Baillière, 1894.

CAREY, N. Factors in the mental processes of school children. *British Journal of Psychology,* 1914, **7,** 453–490.

CORBALLIS, M. C. Rehearsal and decay in immediate recall of visually and aurally presented items. *Canadian Journal of Psychology,* 1966, **20,** 43–51.

DAY, W. F., & BEACH, B. R. *A survey of the research literature comparing the visual and auditory presentation of information.* (AF Technical Report No. 5921) Dayton, Ohio: Wright-Patterson Air Force Base, 1950.

GATES, A. I. Correlations and sex differences in memory and substitution. *University of California Publications in Psychology,* 1916, **1,** 343–350. (a)

GATES, A. I. The mnemonic span for visual and auditory digits. *Journal of Experimental Psychology,* 1916, **1,** 393–404. (b)

GATES, A. I. Variations in efficiency during the day, together with practice effects, sex differences, and correlations. *University of California Publications in Psychology,* 1916, **2,** 1–56. (c)

HAO, Y. T. The memory span of 600 Chinese school children in San Francisco. *School and Society,* 1924, **20,** 507–510.

HAWKINS, C. J. Experiments on memory types. *Psychological Review,* 1897, **4,** 289–294.

HENMON, V. A. C. The relation between mode of presentation and retention. *Psychological Review,* 1912, **19,** 79–96.

JENSEN, A. R., COLLINS, C. C., & VREELAND, R. W. A multiple S-R apparatus for human learning. *American Journal of Psychology,* 1962, **75,** 470–476.

KAY, B. R. Intra-individual differences in sensory channel preference. *Journal of Applied Psychology,* 1958, **42,** 166–167.

KIRKPATRICK, E. A. An experimental study of memory. *Psychological Review,* 1894, **1,** 602–609.

KITSON, H. D. The scientific study of the college student. *Psychological Monographs,* 1917, **23**(1, Whole No. 98).

KOCH, H. L. Some factors affecting the relative efficiency of certain modes of presenting material for memorization. *American Journal of Psychology,* 1930, **42,** 370–388.

MACDOUGALL, R. Recognition and recall. *Journal of Philosophy, Psychology, and Scientific Method,* 1904, **1,** 229–233.

MCGHIE, A., CHAPMAN, J., & LAWSON, J. S. Changes in immediate memory with age. *British Journal of Psychology,* 1965, **56,** 69–75.

MEYER, V. Psychological effects of brain damage. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology.* New York: Basic Books, 1961.

MÜNSTERBERG, H., & BINGHAM, J. Memory. *Psychological Review,* 1894, **1,** 34–38.

O'BRIEN, A. A quantitative investigation of the effect of mode of presentation upon the process of learning. *American Journal of Psychology,* 1921, **32,** 249–283.

SPERLING, G. A model for visual memory tasks. *Human Factors,* 1963, **5,** 19–31.

WAUGH, N. C. Serial position and the memory span. *American Journal of Psychology,* 1960, **73,** 68–79.

WECHSLER, D. *The measurement and appraisal of adult intelligence.* (4th ed.) Baltimore, Md.: Williams & Wilkins, 1958.

WISSLER, C. The correlation of mental and physical tests. *Psychological Review Monograph Supplement,* 1901, **3**(6, Whole No. 16).

WORCHESTER, D. A. Memory by visual and auditory presentation. *Journal of Educational Psychology,* 1925, **16,** 18–27.

(Received January 11, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 2, 132–137

SIMULTANEOUS AND SUCCESSIVE DISCRIMINATION OF SIMILAR LETTERS

JOANNA P. WILLIAMS[1] AND MARGARET D. ACKERMAN
University of Pennsylvania

The first experiment involved 40 first-grade children in a comparison of simultaneous and successive discrimination tasks under varying degrees of stimulus similarity, that is, *b* versus *d*, or *s* versus *b*. In the second experiment, using 48 first graders, simultaneous and successive presentation of highly similar letters (*b* and *d*) were compared in a paired-associate paradigm that simulated the early reading task. Results indicated that first graders learn to discriminate and to respond to very similar letters that are reversals of one another more easily when the letters are presented successively. Suggestions for the development of training programs in early reading are made.

It is generally acknowledged that the letters of the alphabet vary in distinctiveness for beginning readers. Investigations have shown that similar letters such as *b* and *d*, and *n* and *u*, are more often confused than are less similar letters such as *a* and *k* (Popp, 1964). Gibson, Gibson, Pick, and Osser (1962) studied the confusions made in a matching-to-sample task using letter-like forms and transformations of these forms. The letter-like forms were designed to incorporate the critical features that distinguish real letters. In this study, visual discrimination of the forms improved from ages 4 to 8. The authors asserted that children used the habits learned in object discrimination as the basis for discriminating the graphic forms. Topological transformations, for example, are critical in discriminating objects, and even the youngest subjects showed few errors on this feature. Other transformations, such as rotations and reversals, are not critical for object discrimination but are, of course, for the discrimination of graphic forms. In this case, there were many errors at age 4, but very few at age 8. The older children had learned that these features were critical for differentiating letters.

Although there is no doubt about the varying distinctiveness of letters, there is little data as to what methods of training are most effective for developing letter discrimination. These data would be of particular value in developing readiness materials, which so far have been developed largely on an intuitive basis.

The purpose of the present study was to explore the relative effectiveness of simultaneous and successive presentation. Two experiments are reported. The first experiment compared simultaneous and successive discrimination tasks under varying degrees of stimulus similarity. In the second experiment, simultaneous and successive presentations of highly similar letters were compared in a paired-associates paradigm that simulated the early reading task.

EXPERIMENT I

In a simultaneous discrimination problem, two or more stimuli are presented at the same time. On every trial, the subject is reinforced for responding to the same one of these stimuli, regardless of its spatial position. In a successive discrimination problem, on the other hand, only one stimulus is presented on each trial. The subject learns to make a differential response to each stimulus. For example, when Stimulus A is presented, pressing the left of two response buttons may be the reinforced response; when Stimulus B is presented, the right button is reinforced.

According to Spence (1952) and Spiker (1963), the successive problem should be more difficult than the simultaneous one because the stimulus compounds appearing on each trial are more similar in the successive problem. In general, empirical research tends toward the indication that given an equal amount of training, performance on a successive problem is in fact poorer than that on a simultaneous problem. However, one cannot be confident of obtaining this result in all experimental situations. It may depend, for example, on how close the locus of response is to the source of the stimuli (Lipsitt, 1961). Generally speaking, the differences between the two training methods, when they appear, tend to be quite small.

---

[1] Requests for reprints should be sent to Joanna P. Williams, Graduate School of Education, University of Pennsylvania, Philadelphia, Pennsylvania.

Loess and Duncan (1952) compared the two discrimination tasks as a function of stimulus similarity. With easily discriminable stimuli, there was no difference between the two types of training. However, a difference did appear—in favor of simultaneous training—when more similar stimuli were presented. That is, the difficulty of the successive problem, relative to that of the simultaneous, increases with greater stimulus similarity. The result can be derived from Spiker's theoretical formulation. Spiker also predicts that the difficulty of both problems will increase as similarity of the stimuli becomes greater.

Again, some experiments have shown these predicted effects, while others have not, as in Lipsitt's (1961) series of studies in which the stimuli were colored lights. Taking the several experiments together, no unequivocal conclusions can be made.

The purpose of this experiment was to investigate the relative difficulty of simultaneous and successive discrimination tasks under varying degrees of stimulus similarity, with letters serving as stimuli. Presumably, Lipsitt's subjects, fourth graders, would have had distinctive and well-learned labels for each of the colors that were used as stimuli in his experiment. This might have attenuated the effects of stimulus similarity. In the present experiment, younger children, who could not easily label the stimuli, were run.

Method

Subjects. Forty first-grade children from an urban school were used as subjects. The experiment was conducted during the spring semester.

Materials. The similar stimuli were the lowercase letters *b* and *d;* the dissimilar stimuli were *s* and *b.* These letters were chosen on the basis of results such as those of Gibson et al. (1962) and Popp (1964) who found in a matching-to-sample task that children of this general age level often confused *b* and *d* and never confused *s* and *b*.

The subject and the experimenter sat on opposite sides of a small table. A white 5 × 8 inch card was secured in position on the table in front of the subject. This was the response card; on it were drawn two circles. The stimuli were presented on similar cards. On each card, two lower-case block letters were printed. On each trial, a stimulus card was placed above the response cards so that each of the circles on the response card was directly underneath one of the two stimulus letters. The subject responded by placing his finger on one of the two circles. As reinforcement, the experimenter dropped a marble into a plastic cup, placed out of the subject's view so that he could hear the marble drop but was not distracted from his task.

Procedure. The subjects were randomly assigned to one of four groups. Half of the subjects were given a simultaneous discrimination problem, and the other half a successive discrimination problem. Within each of these groups, half of the subjects were trained with highly similar stimuli, and the other half with dissimilar stimuli. The subjects were run individually.

The same instructions were given to all subjects. They were told that they were going to play a game with cards and letters. When a card was presented, they were to look at it and then touch the circle that they thought was correct. They were shown the marbles which served as reinforcements and were told that if they earned enough marbles, they could trade them for a toy after they finished playing the game. In simultaneous training, the subject was presented with a card on which both letters (e.g., *b* and *d*) appeared. Half of the subjects were reinforced for choosing *b*, that is, for putting their finger on the circle below *b;* and the other half, for *d*. A noncorrection procedure was followed. The order of the stimulus arrangements on the cards followed a Gellerman series. In the successive problem, the subject was presented with a card containing one of the two letters, printed twice. The subject was required to make a spatial response; that is, when *d* was presented, the left circle was reinforced. Of course, for half the subjects, the correct response for *b* would be the left circle.

Training continued until the subject reached a criterion of nine consecutive correct responses, or for a total of 54 trials.

Results

Two response measures were used. The first was the number of trials taken to reach criterion. For those subjects who failed to reach the criterion, a score of 54, the total number of trials, was assigned. Table 1 shows the mean number of trials to criterion for each group. Analysis of variance indicated that method of training (simultaneous or successive) was not a significant factor ($F < 1$, $df = 1/36$), nor was stimulus type ($F = 1.76$, $df = 1/36$). The interaction, however, was significant ($F = 4.39$, $df =$

TABLE 1
MEAN NUMBER OF TRIALS TO CRITERION (EXPERIMENT I)

Stimuli	Training		
	Simultaneous	Successive	Total
Similar (b-d)	41.1	29.5	70.6
Dissimilar (s-b)	37.4	46.0	83.4
Total	78.5	75.5	

1/36, $p < .05$). With highly similar stimuli, *b* and *d*, the successive problem was less difficult than the simultaneous problem, contrary to expectations, while with the dissimilar stimuli, *s* and *b*, the successive problem was more difficult.

The second response measure was the number of correct responses made during the last 18 trials, the final third of training. Table 2 shows the mean for each group. Again, neither main effect was significant (training method: $F < 1$, $df = 1/36$; stimulus type: $F < 1$, $df = 1/36$), and the interaction was significant ($F = 7.92$, $df = 1/36$, $p < .01$).

A substantial proportion of the subjects (42%) failed to reach criterion before the experiment was terminated (six in the simultaneous-similar condition, two in the successive-similar, three in the simultaneous-dissimilar and six in the successive-dissimilar). This figure is not uncharacteristic of experiments of this type.

Discussion

The results of this experiment are in direct contradiction to predictions from Spiker's theory, and they also differ from some empirical findings. The present experimental situation appears to have been comparable to those of other studies. However, although there is no doubt that the stimulus pair *b* and *d* qualify as more similar than the pair *s* and *b*, these letter stimuli may well differ in an important way from the stimuli used previously.

TABLE 2
MEAN NUMBER OF CORRECT RESPONSES OVER LAST 18 TRIALS (EXPERIMENT I)

Stimuli	Training		
	Simultaneous	Successive	Total
Similar (b-d)	12.5	16.6	29.1
Dissimilar (b-s)	16.2	13.3	29.5
Total	28.7	29.9	

Most studies have utilized stimuli such as colors or shades of gray, all of which can be identified and labeled differentially by the subject. In fact, Loess and Duncan (1952), who worked with college students, defined stimulus difficulty in terms of the number of cues the subject had to discriminate in order to choose the positive stimulus consistently. Their stimuli were of the type used in concept-attainment tasks, for instance, in one task "circle on white, square on gray" was correct as opposed to "circle on gray, square on white."

In the present experiment, however, it is likely that the differences between *b* and *d* could not be defined by the subjects because *b* and *d* are notoriously confusing for many first graders. Thus whatever cues were to be used in the solution of the discrimination problem were not easily available to the subject, and in fact, had to be developed during the course of training.

On the other hand, for the *s-b* comparison, distinctive cues are easily identified; for example, *s* is a smaller letter and comparisons in terms of size are, of course, made by children of this age with no difficulty.

According to Gibson and Gibson (1955), improvement in discrimination comes about through the learning of distinctive features of the stimuli to be discriminated. During practice, the subject responds to an increasing number of features, and he discovers which ones do indeed distinguish between the two stimuli.

What is suggested here is that the subject who is faced with a *s-b* problem will find it easy to identify a critical feature—size, for example. He can then use this feature in learning which of the two stimuli is correct in the discrimination task. In this situation, as when colors are used as stimuli, simultaneous training will be superior to successive training as predicted by Spiker's theory.

However, *b* and *d* are stimuli for which distinctive features cannot easily be detected. If a pair of stimuli such as *b* and *d* are presented simultaneously, then there is so much information presented during a single trial that the subject cannot focus as easily on any particular feature in order to develop the basis on which he would be able to make the differentiation. It appears that successive discrimination training, in which the subject must attend to only one stimulus at a time, will aid the identification of critical features, which then can be used in comparison in solving the problem.

These findings are similar to those of a study of cross-modality (visual-haptic) recognition of shape (Lavrent'eva & Ruzskaia, 1960) in which successive comparison tended to be superior to simultaneous comparison. These authors suggested that when the subject compared the standard and the matching figures at the same time, interference was produced, which led to recognition errors. It is probably significant that in their study, as in the present one, the task was rather difficult.

EXPERIMENT II

Because the results of the first experiment were so different from what was expected, we decided to see if the same result would occur in a more complex task. In this experiment, highly similar letters (*b* and *d*) were combined into trigrams which were presented in a paired-associate paradigm. This task was chosen because it simulated the early reading task, in which the child must learn to associate phonemes with their graphic values.

Two combinations were possible. First, the stimuli could be presented simultaneously in a single trigram—*bad* (or *dab*). This approximates the simultaneous training studied in Experiment I, in which both *b* and *d* appeared on every trial. Or, on the other hand, the stimuli could be presented successively, that is, with only one of the two letters appearing in a given trigram—*bab* (or *dad*). This approximates the successive training of Experiment I.

This is an unconventional use of the terms "simultaneous presentation" and "successive presentation," but it is clear that the visual configuration present in each trigram is comparable to that present in the corresponding discrimination task. In light of the results of Experiment I, it was predicted that performance would be superior when the highly similar stimuli were presented successively, that is, when the trigrams were *bab* or *dad* and not *bad* or *dab*.

Method

Subjects. Forty-eight first-grade children, chosen from the top four of six reading groups in an urban school, were used as subjects. The experiment was conducted during the spring semester, and none of the subjects had participated in Experiment I.

Materials. Each list consisted of four trigrams. In addition to the two critical trigrams in each list, there were two additional trigrams, termed the noncritical stimuli. These were either *ban* and *sad*, or *dan* and *sab*. These trigrams were chosen so that the critical trigrams could not be differentiated on the basis of either the first or last letter alone. Pilot work indicated that a list that included all four noncritical trigrams was too difficult a task. It was decided, therefore, to use two sets of two noncritical trigrams, to counterbalance the effects of specific letters. Half the subjects in each condition (simultaneous presentation and successive presentation) were given one set of noncritical trigrams, and the other half learned the other set.

The trigrams were paired with colors as response terms. There were two combinations of trigram-color pairings in order to control for any cues specific to any particular pair.

The physical setting was the same as in Experiment I. The trigram stimuli were printed in lowercase block letters and the colors were placed above them. The subjects were run individually.

Procedure. The subjects were randomly assigned to either the simultaneous presentation or the successive presentation group. The anticipation method with correction was used. The subject was allowed approximately 5 seconds in which to respond.

Training continued until there were no errors on two consecutive trials, a trial being defined as presentation of all four stimulus pairs, or until nine trials had been given.

Results

Two response measures were used. The first was the number of correct responses and the second was the number of confusion errors made on the critical trigrams. In the analyses, the subgroups were combined to form two groups, simultaneous presentation and successive presentation.

Table 3 presents the mean number of correct responses for the two groups on each trigram. Orthogonal *t* test comparisons were selected to analyze the differences between

TABLE 3

MEAN NUMBER OF CORRECT RESPONSES (EXPERIMENT II)

Training	Stimuli			
	Critical		Noncritical	
Simultaneous	2.58 (dab)	5.04 (bad)	4.71 (ban-dan)	6.17 (sad-sab)
Successive	3.79 (dad)	6.12 (bab)	5.96 (ban-dan)	6.08 (sad-sab)

trigrams. First, a comparison between the critical trigrams in the simultaneous condition and those in the successive condition indicated that the successive critical trigrams were easier to learn than the simultaneous critical trigrams ($t = 2.08$, $df = 184$, $p < .05$), as had been predicted. Other comparisons were also made. The comparison between the critical trigrams starting with *b* (*bad* and *bab*) and the critical trigrams starting with *d* indicated that the initial *d* trigrams were more difficult than were initial *b* trigrams ($t = 6.26$, $df = 184$, $p < .001$). Comparison of the noncritical trigrams in the simultaneous presentation versus successive presentation conditions showed that they did not differ in difficulty ($t = 1.04$, $df = 184$), which indicates that the noncritical trigrams did not affect performance on the critical trigrams. The comparison between the critical and the noncritical trigrams indicated that the critical trigrams were more difficult ($t = 3.55$, $df = 184$, $p < .01$). And finally, the noncritical trigrams *ban-dan* did not differ in difficulty from the noncritical trigrams *sad-sab* ($t = 1.44$, $df = 184$).

TABLE 4

NUMBER OF CONFUSION ERRORS–CRITICAL TRIGRAMS (EXPERIMENT II)

Training	Class of confusion error			Total	χ^2[a]
	I	II	III		
Simultaneous					
dab	112 (bad)	26 (ban-dan)	12 (sad-sab)	150	117.[illegible]8**
bad	57 (dab)	22 (ban-dan)	9 (sad-sab)	88	42.02**
Successive					
dad	73 (bab)	24 (ban-dan)	22 (sad-sab)	119	42.06**
bab	27 (dad)	23 (ban-dan)	13 (sad-sab)	63	4.95*

[a] $df = 2$.
* $p < .10$.
** $p < .001$.

The number of confusion errors made on the critical trigrams provided a second response measure. A confusion error consisted of a response that was incorrect for the particular trigram responded to, but was correct for another trigram in the list. Table 4 presents the number of confusion errors for the critical trigrams. The types of confusion errors are classified under three categories: Type I errors, sharing both initial and terminal configuration, that is, the (color) response correct for *dad* is given to *bab;* Type II, sharing initial configuration, that is, the response to *ban* is given to *dad;* and Type III, sharing terminal configuration, that is, the response for *sad* is given to *dad.*

Chi-square analyses of the types of errors were made for each trigram. The hypothesis tested in each case was that the confusion errors would be equally distributed among the three error types. The chi-square values and the probability levels are included in Table 4, and it can be seen that the errors were not equally distributed. In each case, it is evident that the largest proportion of errors were of Type I. That is, the critical trigrams were more often confused with the trigram that shared both initial and terminal configuration—the other critical trigram.

Discussion

This was a difficult task, as shown by the number of confusion errors. Moreover, only 2 out of 24 subjects in the simultaneous presentation condition reached criterion, whereas 6 out of 24 subjects in successive presentation reached criterion. However, the results supported the prediction. That is, successive presentation of the similar stimuli *b* and *d* resulted in reliably better learning in a paired-associate task, just as successive presentation had proved superior in the discrimination task of Experiment I.

It is interesting to note that the critical trigrams starting with *d* were more difficult than the critical trigrams starting with *b.* It may be that children develop effective discrimination of the letter *b* earlier than they do with *d.*

Sad and *sab* could be discriminated on the basis of the initial letter alone, since only one of the two was presented to each subject and it was the only trigram that began with a smaller letter. For the same reasons, *ban* and *dan* could be discriminated on the basis of the terminal letter only. It appears that the subjects did tend to use this information, in that there were significantly more correct responses on noncritical trigrams than on critical trigrams. However, results such as those of Williams, Blumberg, and Williams (1970) which show that first-grade children recognize a word on the basis of its first letter would predict that *sad* or *sab* would be learned more quickly than *ban* or *dan*. Such was not the case in this experiment; the two sets of noncritical trigrams were learned equally well. It is possible that the subjects did not use the method of attack that they used in the rather different tasks presented by Williams et al. (1970).

The results of the two experiments taken together seem to indicate that first-grade children learn to discriminate and to respond to very similar letters that are reversals of one another more easily when the letters are presented successively than when they are presented simultaneously. This finding held true for both a visual discrimination task and for a paired-associate task.

Samuels (1969) compared what he termed simultaneous and successive discrimination training with kindergarten children. The training was different from the tasks in the experiments presented here. It was in some ways similar to a matching-to-sample and delayed matching-to-sample procedure. The subjects were kindergarten children and the stimuli were *p, q, b,* and *d*. The author concluded that subjects learned the matching task faster under simultaneous training. However, subjects who had had the successive training did better on a subsequent paired-associate task involving the same stimuli. Thus both Samuels' experiment and the present one indicate that the presentation of only one of the highly similar stimuli at a time leads to more efficient learning of paired associates.

In conclusion, successive presentation of highly similar letters appears to result in both better visual discrimination and in better paired-associate learning of tasks that incorporate these letters. Our hypothesis is that the subject seems to be able to distinguish more efficiently a critical feature when only one stimulus is presented at a time. With less similar stimuli, it is possible that simultaneous training would be more efficient because it would not be as difficult for a critical feature to be identified.

These results suggest that readiness programs that include letter discrimination and letter naming training can be most effective if different methods are utilized for different letters. Letters which are distinctive can be introduced simultaneously, and they will probably require less training than will letters that are highly similar; these should be introduced successively.

REFERENCES

GIBSON, E. J., GIBSON, J. J., PICK, A. D., & OSSER, H. A developmental study of the discrimination of letter-like forms. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 897–906.

GIBSON, J. J., & GIBSON, E. J. Perceptual learning: Differentiation or enrichment? *Psychological Review*, 1955, **62**, 32–41.

LAVRENT'EVA, F. V., & RUZSKAIA, A. G. Comparative analysis of touch and vision. *Soviet Psychology and Psychiatry*, 1960, **1**, 28–31.

LIPSITT, L. P. Simultaneous and successive discrimination learning in children. *Child Development*, 1961, **32**, 337–347.

LOESS, H. B., & DUNCAN, C. P. Human discrimination learning with simultaneous and successive presentation of stimuli. *Journal of Experimental Psychology*, 1952, **44**, 215–221.

POPP, H. M. Visual discrimination of alphabet letters. *The Reading Teacher*, 1964, **17**, 221–226.

SAMUELS, S. J. Effect of simultaneous versus successive discrimination training on paired-associate learning. *Journal of Educational Psychology*, 1969, **60**, 46–48.

SPENCE, K. W. The nature of the response in discrimination learning. *Psychological Review*, 1952, **59**, 89–93.

SPIKER, C. C. The hypothesis of stimulus interaction and an explanation of stimulus compounding. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior*. Vol. 1. New York: Academic Press, 1963.

WILLIAMS, J. P., BLUMBERG, E. L., & WILLIAMS, D. V. Cues used in visual word recognition. *Journal of Educational Psychology*, 1970, **61**, 310–315.

(Received January 11, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 2, 138–140

WORD SCANNING RATE FOR GOOD AND POOR READERS[1] †

LEONARD KATZ[2] AND DAVID A. WICKLUND
University of Connecticut

Twenty fifth-grade poor readers and 19 good readers were presented with a series of trials on each of which two successive slides were shown. The first slide contained a single word. The second slide contained either two or three words which comprised either (*a*) a grammatical, meaningful sentence or (*b*) a scrambled version of a grammatical sentence. The subjects were instructed to respond "yes" or "no" depending on whether the first slide word did or did not appear on the second slide. Although poor readers were about 250 milliseconds slower overall than good readers, both groups had equivalent scanning rates. Reaction time to the three-word sentences was about 100 milliseconds slower than reaction time to the two-word sentences, for both groups. The absence of effects due to grammaticality and the absence of an interaction between sentence length and "yes" versus "no" suggested that subjects in both groups used a rapid exhaustive perceptual scan. The differences observed between good and poor readers cannot be caused by differences in the transformation and matching process.

Consider the reading of a simple sentence to be composed of the following sequence of processes. Initially, visual information (part of a word, a word, or several words) is received and encoded as an essentially visual abstraction, for example, letter and word forms (cf. Sternberg, 1967). Then, a phonemic representation is generated and the visual information either decays or is not further utilized (Posner, Boies, Eichelman, & Taylor, 1969). The phonemic representation interacts with memory to produce meaning (Norman, 1968). The differences between good and poor readers in reading speed may exist at any or all points of the sequences: in the initial visual read-in, in the phonemic transformation, or in the memory search.

The present experiment attempted to locate the sources of differences between good and poor readers more precisely. On each trial, the subjects were presented with a sentence to read and then asked to report whether or not a predetermined key word had occurred in the sentence. The sentences were either two or three words in length. The time from the onset of the sentence to the report (either "yes" or "no") was measured.

To accomplish the task, the subject must receive visual information from the sentence, transform it, match the transformed information to the memory of the key word, choose the appropriate response and, finally, produce the motor activity necessary to effect a response. The time required for read-in, transformation, and matching can be separated from the total response time; what is needed is the relation of reaction time as a function of the number of words in the probe sentence. The slope of the function reveals the time required for scanning, transformation, and matching; the intercept of the function reveals the time required for orienting to the probe, choosing a response (response decoding), and producing the response. The technique is similar to that used by Briggs and Blaha (1969). Thus, if poor readers differ from good readers in the slope of the reaction time function, we know that this is due to differences in the word scanning, transformation, and/or matching processes. If the differences between poor and good readers are in the intercept of the reaction time function we can infer that they differ in the ability orienting to the probe, choosing a response, and/or producing a response. A third possibility is that poor and

---

[1] Supported by Grant HD-03932 from the National Institute of Child Health and Human Development. Linda Lesniak and Michael Dorman ran the subjects. The authors are indebted to Carol Hill, Daisy Pilcher, and Nathan Joy of the Manchester, Connecticut, school system.

[2] Requests for reprints should be sent to Leonard Katz, Department of Psychology, University of Connecticut, Storrs, Connecticut 06268.

good readers differ in both slope and intercept.

The present study also attempted to study differences in processing time between good and poor readers as a function of the grammaticality of the probe sentence. If good readers are more sensitive to the grammar of the probe sentence than poor readers, then presumably, their scanning rate should be faster. With nongrammatical sentences, this difference should be attenuated.

METHOD

Stimuli

Twenty grammatical, meaningful probe sentences in each of three lengths (two, three, and five words) were constructed.[3] Words were chosen from common words within the vocabulary of most fifth graders. Within each set of a given length, each serial position within the sentences was chosen equally often to be a true target word, that is, a word which appeared in the probe sentence. For every true target word, a false target word was chosen which began with the same letter as the true probe and was approximately the same length as the true target word. Thus, there were 40 sentences in each of the three lengths for which 20 true target words and 20 false target words were appropriate.

Two lists of 20 sentences each were constructed from the 40 sentences in each length (labeled the A and B lists) such that each list contained 10 true and 10 false targets. Each sentence appeared once in each list; if a given sentence had a true target in List A it was given a false target in List B and vice versa.

Each of the 40 grammatical sentences was scrambled to produce a nongrammatical sentence. The scrambling was done so that all serial positions were chosen equally often as target items.

Subjects

Forty fifth-grade subjects were chosen and were designated as good or poor readers on the basis of Iowa reading scores. The range of good readers was 54 to 97 with a median score of 58, and the range of poor readers was 1 to 38 with a median score of 25. Ten good readers and 10 poor readers were run under the grammatical sentence condition; the remaining 10 good readers and 10 poor readers were run under the nongrammatical sentence condition. Within each group of 10, half the subjects were run with the A set and half with the B set. One good reader had to be eliminated due to experimenter error, leaving a total *N* of 39.

Procedure

Each subject went through a practice list of Sentence Length 4 followed by three lists of Lengths 2, 3, and 5. The order of presentation of lengths was balanced so that Length 2 preceded Length 3 five times and followed five times in each experimental group; Length 5 was presented first four times, second four times, and third twice in each group.

Stimuli were presented via a Lafayette KT-800 projecting tachistoscope. The first stimulus word (the target) was presented for 5 seconds and after a 900 millisecond slide-change interval, was followed by the probe sentence which remained on for 5 seconds. A Hunter Model 1520 clock was started by the presentation of the probe sentence. The subject responded "yes" if the probe sentence contained the target word and "no" if it did not. A voice-operated relay stopped the clock.

RESULTS AND DISCUSSION

Errors for all subjects were few (less than 5%) and unsystematic. Inspection of the data showed that the mean latency in each condition tended to be inflated by very large latencies (e.g., 4 or 5 seconds) that were occasionally produced by some subjects, most often by poor readers. These long latencies probably reflected processes not of interest to the present study. For example, if the subject forgot the target stimulus (but

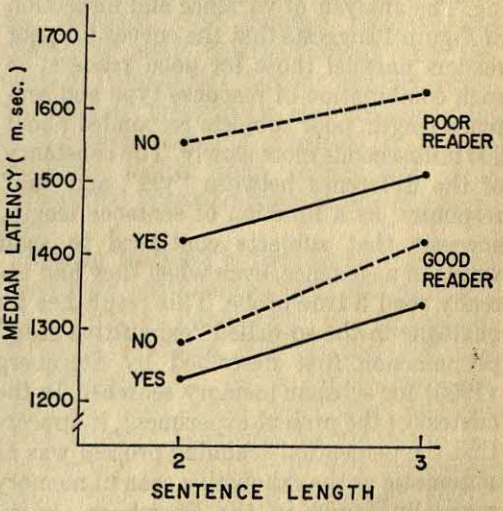


FIG. 1. Latency of response as a function of sentence length, reader ability, and type of response.

---

[3] Due to experimenter error, several subjects received improper stimulus sequences under Length 5. For the latter reason and because interactions of Length 5 and its ordinal position of presentation were observed, the data for Length 5 were not included in the analysis.

guessed correctly) a longer latency than usual may have been produced. Therefore, the median latency for each subject in each condition was chosen as the dependent variable because it was believed to represent more accurately the central tendency of the reaction time distribution for true correct responses.

An analysis of variance indicated significant effects due to sentence length, $F = 13.7$, $df = 1/35, p < .01$; reader ability, $F = 4.2$, $df = 1/35, p < .05$, and response type (yes versus no), $F = 9.1$, $df = 1/35$, $p < .01$. Grammaticality did not approach significance nor did any of its interactions.

The lack of any effect of grammaticality suggests that the subjects may not have read the sentences for meaning. Instead, the subjects may have simply scanned both grammatical and nongrammatical sentences at a rapid rate, treating each word in the sentence independently of other words. Thus, lack of an effect of grammaticality or lack of an interaction between grammaticality and reader ability may suggest only that the present experiment was not the appropriate one to test the grammaticality hypotheses.

Figure 1 presents median latency as a function of sentence length for good and poor readers and responses "yes" and "no." Both good and poor readers produced median latencies that indicate rapid responding. The analysis of variance and inspection of Figure 1 suggests that the curves for poor readers parallel those for good readers; in each combination of response type and sentence length, poor subjects responded about 200 milliseconds more slowly. The constancy of the difference between "yes" and "no" responses, as a function of sentence length, suggests that subjects continued to scan words in a sentence, even when they had already read a true probe. This result has its analogue in the so-called "exhaustive scan" phenomenon first described by Sternberg (1966) for subspan memory searches. In the context of the present experiment, it appears that the perceptual scanning process was as automatic as the exhaustive scan of memory generally found in the Sternberg experiments. It may be the case that it is more efficient for subjects to utilize an exhaustive scanning procedure, one that has a preprogrammed stopping point, than to utilize a scan that must be prepared to terminate at a point undetermined at the beginning of the scan. It should be noted that this argument applies equally well whether the scanning process is serial or parallel in nature. Recently, Atkinson, Holmgren, and Juola (1969) found exhaustive scanning in a visual scanning experiment with adult subjects. The lack of a grammaticality effect also suggests a rapid word-oriented scan that literally does not take the time to consider syntactic relations among words.

Figure 1 shows that both good and poor readers took about 100 milliseconds longer to scan a three-word sentence than a two-word sentence. The absence of a significant interaction between sentence length and reader ability, together with a significant effect of reader ability, suggest that good and poor readers differ in the intercept of their reaction time functions but not in the slopes of these functions. This indicates that the good and poor readers do not differ in the ability to scan, transform, and match words; the differences between the two types of readers occur somewhere else in the chain: in the orientation to the probe, in response decoding, or in the motor portion of the response.

REFERENCES

ATKINSON, R. C., HOLMGREN, J. E., & JUOLA, J. F. Processing time as influenced by the number of elements in a visual display. *Perception and Psychophysics*, 1969, **6**, 321–326.

BRIGGS, G. E., & BLAHA, J. Memory retrieval and central comparison times in information processing. *Journal of Experimental Psychology*, 1969, **79**, 395–402.

NORMAN, D. Toward a theory of memory and attention. *Psychological Review*, 1968, **75**, 522–536.

POSNER, M. I., BOIES, S. S., EICHELMAN, W. H., & TAYLOR, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology*, 1969, **79**(1, Pt. 2).

STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

STERNBERG, S. Two operations in character recognition: Some evidence from reaction-time measurements. *Perception and Psychophysics*, 1967, **2**, 45–53.

(Received January 26, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 2, 141-147

ORDER EFFECTS IN THE AFFECTIVE LEARNING STYLES OF OVERACHIEVERS AND UNDERACHIEVERS†

JOSEPH F. RYCHLAK[1]
Purdue University

AND

THOMAS J. TOBIN
Rockhurst High School, Kansas City

The predilection to learn positively rated trigram materials more efficiently than negatively rated materials was put to test in a sample of 64 high school males identified via grade point average and IQ as overachievers, underachievers, above average, and below average students. All subjects learned two lists of paired-associate trigrams, one of which had been rated as liked by the subject and one which he had rated as disliked (unmixed lists). In line with experimental predictions, the overachiever showed the least disparity between positive and negative reinforcement value, and the underachiever showed the greatest reinforcement value effect. Learning disliked lists before liked led to uniformly positive nonspecific transfer for all subjects, whereas learning liked before disliked materials did not facilitate transfer this clearly.

It has been shown in several paired-associate studies that affective evaluations of consonant-vowel-consonant "trigrams" can influence rate of acquisition independently of the typical measures of association value (Abramson, Tasto, & Rychlak, 1969; Rychlak, 1966; Tenbrunsel, Nishball, & Rychlak, 1968). Association value is typically measured by asking the subject whether or not a trigram can be recognized as a word, or as having word qualities such as connotations or other "associations" (e.g., Archer, 1960). In an affective measure the experimenter asks the subject to judge whether trigrams are "liked versus disliked," and then, while controlling for possible effects of association value, demonstrates a facilitating effect for such uniquely determined emotive ratings in subsequent learning tasks. Although other terms might be used to name this idiographic preference, such as "affective significance" or "interest value," the most general and abstract designation seems to be that of "reinforcement value." To date, the typical finding for normal subjects has been that positive reinforcement value (material judged as "liked") facilitates learning in relation to negative reinforcement value (material judged as "dislike").

The theory underlying this area of study has it that association value measures are a function of frequency (incidence, associative ties, probability in a language structure, etc.) whereas reinforcement value measures are—though not unrelated to frequency factors in a correlational sense—independent assessments which the subject makes of learning materials. Considered in terms of an item's meaningfulness (i.e., significance or import which it holds for the subject), association value is seen as taking the form of a quantitative, continuous metric of more-to-less, whereas reinforcement value is held to be a qualitative, dichotomous judgment of either-or. In short, this view holds that the subject is constantly coming up onto experience with an evaluating intellect. The number (quantitative metric) of input experiences and other frequency (countable) factors like rehearsal, warming up, recognition, etc., are important to the acquiring and retention of certain meanings rather than others, but it is argued that a second basic factor is also at play.

Human learning is thus to be viewed as receiving a contribution from the individual's affective evaluation of potentially learnable materials. It is not that the subject evaluates like versus dislike and then rehearses liked items more than disliked items, thereby acquiring his ability to retain entirely on the basis of such frequency factors (the number of rehearsed trials, "connections," or mediations). Rather, it is that the subject tautologically identifies himself with items in the environment. It is this tautological identification of a "liked (self)" with a "liked (item)" which is now said to be at play in addition to the usual frequency factors already mentioned. This explanation offers an alternative and additional way of thinking about human learning with frequency factors partialed out of the account. Of course, any series of data collections by way of experimentation can be explained theoretically in literally *N* potential ways (Rychlak, 1968). Hence, one does not have to accept this theoretical suggestion in order to accept the empirical fact that reinforcement value is at play in human learning independently of association value.

---

[1] Requests for reprints should be sent to Joseph F. Rychlak, Department of Psychology, Purdue University, Lafayette, Indiana 47907.

It seemed reasonable to expect maximal differential effects in learning due to affectivity among overachievers and underachievers. Underachievers have been repeatedly found to be highly emotional individuals (Dowd, 1952; Holland, 1959; Lum, 1960; Mitchell, 1959). The parents of overachievers are likely to offer advice and encouragement to their children, whereas parents of underachievers are found to be punitive (Ringness, 1965). The self-image which results is significant in that overachievers rate themselves more favorably than they rate their peers, whereas underachievers rate themselves less favorably than they rate their peers (Peppin, 1963). The overachiever is more likely to be considered emotionally mature than is the underachiever (Cohler, 1940; Horrall, 1957; Morgan, 1952).

The investigators theorized that the positive versus negative reinforcement value effects to be noted among an unselected sample of subjects should be increased when measured in a group of underachievers. As emotional individuals, highly likely to be swayed by how they "feel" about the learning task confronting them, underachievers might be expected to show a greater disparity between positive and negative reinforcement value factors than other individuals. On the other hand, overachievers, who have exerted considerable effort to rise above their expected level of performance, should be more likely to show an eclipsed disparity between liked and disliked materials. The overachiever tries to overcome his natural tendency to reject material he dislikes, and hence, even more so than the bright individual functioning at his above-average level, the overachiever should reflect a diminishment in the reinforcement value effect. The mediocre individual, functioning at an average or even below average level in school might be expected to have a fairly marked disparity between liked and disliked materials, but not quite as great as that found among underachievers. Hence, the predicted rank ordering from low to high "positive reinforcement value effect" in this research was as follows: overachievers, above average students (i.e., bright individuals at their level), below average students (i.e., relatively mediocre students at their level), and underachievers.

In order to circumvent difficulties with the use of a mixed list (combining positive and negative reinforcement value materials in one trigram list) the investigators decided that the most stringent test of the above hypotheses was to be obtained if a subject were asked to learn two unmixed lists. This called for a counterbalancing procedure, but in the process it was noted that a possible order effect could be expected. That is, the subjects would be learning liked before disliked materials and vice versa. This offered a unique opportunity for reinforcement value to be paralleled to earlier work on transfer effects, particularly on what has been termed nonspecific transfer (Kausler, 1966, p. 361). In an A–B, C–D paradigm one expects to find positive transfer effects, usually said as emanating from warm-up or learning-to-learn facilitation across lists (Harlow, 1949; Postman & Schwartz, 1964). Up to a certain point, as the number of previous experiences on lists increases, the subject's performance in the "present" task-list is facilitated. There is also reason to believe that varying the levels of association value meaningfulness (from high to low and vice versa) has a systematic effect on nonspecific positive transfer as practice on successive lists is carried out (Cieutat, Stockwell, & Noble, 1958).

If one now equates number of trials with number of previous lists it seems apparent that a natural prediction to make is, nonspecific transfer effects will be greater when the more difficult material (requiring more trials to learn) is learned first, and vice versa. Hence, by extension, from what is known of reinforcement value effects, it follows that learning liked lists previous to disliked lists should not result in as much general or nonspecific transfer as when the disliked lists are learned first. Since the tautologizing of experience implies that the subject already knows more about his liked materials, it follows that he does not have to devote the trials, hence acquire, the learning-to-learn facets of practice in order to achieve the task at hand. With disliked lists, however, the subject must indeed devote considerable effort to such tactics and these in turn should accrue much more dramatically to the learning of a second, unrelated but liked list.

METHOD

Subjects

The subjects were 64 male high school students, selected from sophomore through senior class levels and ranging in age from 15 to 18 years. These subjects constituted four subgroups of 16 subjects in each of the four conditions under investigation. A subject's assignment to one of these conditions was determined as follows: the three grade levels were first ranked according to a student's cumulative grade point average, and then subdivided into thirds (upper, middle, lower). Since the grade point average tends to vary across the years it was decided to treat the three grade levels separately. The number of students broken into thirds at each year level was as follows: 202 sophomores, 198 juniors, and 183 seniors.

After these subjects had been identified in terms of their level of grade attainment they were reranked on the basis of an IQ measure. The Henmon-Nelson group IQ test was available for all students in this school system. Once again, each subject was identified according to whether he fell in the top, middle, or lower third of the group—this time considered all together. Thus, an IQ of 127 or higher proved to be the range of the upper third, and an IQ of 118 or lower proved to be the upper limit of the bottom third of all subjects. As a group instrument, the Henmon-Nelson has been shown to overestimate the level of IQ one might expect from an individual test such as the Stanford Binet. Cronbach (1960, p. 221) notes that it is not unusual for twice as many subjects to be classified as having IQs in excess of 120 for the Henmon-Nelson as compared to the Binet. However, the overall correlation between these instruments is regularly of the value .70 to .90 and hence the investigators did not feel that the IQ value as such need influence the decision to use the group scores available. Hence, all subjects with IQs of 127 or higher were considered "above average" and those with IQs below 119 were considered "below average."

After each class level had been broken into three divisions on the two dimensions of grade point and IQ, the experimental conditions were filled empirically as follows: A subject in the top third on IQ and grade point average was considered an above average student; if the subject was in the bottom third on both IQ and grade point average he was considered a below average student; anyone with a top third IQ and a lower third grade point average record was termed an underachiever; and conversely, the subject with a bottom third IQ and a top third grade point average was considered an overachiever. These global designations were not important in themselves, of course, because the investigators were trying to establish conditions which might enter into factorial interactions rather than attempting to typologize the subjects. Sample size was limited by the most difficult condition to fill—that is, the overachiever! When all of the possible overachievers had been identified, the rest of the conditions were filled by selecting randomly from the subjects meeting one of the four designations.

Procedure

All subjects were contacted individually by telephone and asked to participate; there were no refusals. The usual two administrations of the Phonetic Preference Inventory, which contains 140 trigrams for rating purposes, were then made (Rychlak, 1966), and paired-associate lists constructed accordingly. Trigrams which have been rated as liked (much, slightly) or disliked (much, slightly) on both administrations are selected. It is also immediately possible to control for association value because the Phonetic Preference Inventory items are arranged numerically so that a distance of a few points on the scale indicates that the trigrams in question are either identical or only a point or two different in nomothetic association value (Archer norms, 1960). Hence, as each liked pair of trigrams is determined, a disliked pair from the same nomothetic association value range is similarly identified.

Eight pairs of trigrams having the characteristic of positive reinforcement value and eight pairs having negative reinforcement value characteristics were arrayed for each subject (necessitating 16 liked and 16 disliked trigrams being taken from the Phonetic Preference Inventory). Thus, each subject was asked to learn two unmixed lists of eight pairs, one consisting of positive and a second consisting of negative reinforcement value trigrams, both of which had been equated for association value. In each of the four experimental groupings, the subjects were randomly assigned to either a "positive-negative" or a "negative-positive" reinforcement value sequence in the learning of their

TABLE 1

Mean and Standard Deviation for Intelligence, Grade Point Average, and Order of Reinforcement Value

Intelligence	Grade point average	Order of lists	Positive RV trigrams		Negative RV trigrams	
			M	*SD*	*M*	*SD*
Above average *IQ*	Upper third	Positive RV first	48.50	17.50	42.38	13.58
		Negative RV first	33.88	11.13	52.00	12.59
	Lower third	Positive RV first	44.75	9.30	48.25	12.95
		Negative RV first	33.25	5.58	69.00	21.43
Below average *IQ*	Upper third	Positive RV first	59.38	10.91	43.13	10.24
		Negative RV first	33.75	5.70	56.00	16.81
	Lower third	Positive RV first	64.25	17.16	57.13	18.17
		Negative RV first	41.38	10.80	66.13	26.08
			44.89	12.70	54.25	16.89

Note.—Abbreviated: RV = reinforcement value. RV main effect, $p < .001$; RV and grade point average interaction, $p < .05$; RV and order interaction, $p < .001$. N for combined sample = 64.

lists. Trigram pairs were flashed onto a screen before the subject by a projector with the rate of approximately 5 seconds per exposure, and 1 second between slides. Between trials the order of the eight pairs was shuffled to prevent serial learning cues. To focus the subject's attention on a given pair, the investigator asked him to state "goes with" following the reading of the first trigram and preceding the appearance of the second (LIX "goes with" TUC). Beginning with the second trial the subject was instructed to call out what the first trigram "goes with" before the second slide made its appearance (method of anticipation). When two correct anticipations had been given a trigram pair was removed from the list, so that the task was one of learning pairs rather than a list as such. Raw data consisted of the total trials taken to learn the eight pairs to a criterion of two consecutive correct anticipations.

RESULTS

Table 1 contains the means and standard deviations of the various conditions of the experiment. A 2^4 factorial analysis of variance was run on these data, with the following results: a significant main effect for reinforcement value ($F = 23.942$, $p < .001$); a significant interaction between reinforcement value and grade point average ($F = 6.454$, $p < .05$); and a truly remarkable interaction between reinforcement value and order of lists ($F = 68.746$, $p < .001$). The expected triple interaction between reinforcement value, grade point average, and IQ—which defines most clearly the overachievers versus underachievers—failed to materialize. Even so, a definite trend can be noted in Table 2, in support of the hypotheses. Table 2 contains the learning data of Table 1, grouped now into the four experimental conditions. Note the difference score column is exactly as predicted, with underachievers obtaining the largest dispar-

TABLE 2

Trials to Criterion Arranged by Experimental Condition

Experimental group	*M*	*SD*	Difference
Underachievers			
Positive reinforcement value	39.00	9.59	19.62
Negative reinforcement value	58.62	20.53	
Below average students			
Positive reinforcement value	52.81	18.34	8.82
Negative reinforcement value	61.63	23.10	
Above average students			
Positive reinforcement value	41.19	16.39	6.00
Negative reinforcement value	47.19	13.95	
Overachievers			
Positive reinforcement value	46.56	15.49	3.00
Negative reinforcement value	49.56	15.33	

Note.—Difference is between positive and negative reinforcement value.

ity between positive and negative reinforcement value, then below average, above average, and finally overachievers, in that order.

It is thus clear that positive reinforcement value has once again been shown to facilitate learning, a facilitation which can be noted across all groups of the study. The reinforcement value effect is noted differentially across grade point average level at the .05 level of significance, with the upper third subjects in grades reflecting a difference score of 4.50 and the lower third subjects reflecting a difference score of 14.22 (based on data from Table 1). There was no significant interaction between IQ and reinforcement value, but it is of interest to note that the upper third IQ subjects reflected a difference score between positive and negative reinforcement values of 12.82, while the lower IQ subjects reflected a difference of 5.91.

Probably the most interesting finding of this study emerges when the order effects are considered. Nonspecific transfer was dramatically greater when the subject was asked to learn his negative reinforcement value trigrams before learning his positive reinforcement value trigrams. Indeed, of the 32 subjects taking the dislike-like sequence all learned their liked lists more quickly than their disliked. The combination of order and reinforcement value had a truly remarkable effect. Of the 32 subjects who learned their liked trigrams before learning their disliked trigrams, 13 still record a more rapid acquiring of the criterion for their liked trigrams. It is significant to note that 10 of these 13 subjects were found among the lowest third in grade point average (six overachievers and four below average subjects). The remaining three of these subjects were above average students. None of the overachievers failed to benefit from the combination of a negative-positive reinforcement value order effect.

Another way of underscoring the facilitation effect of negative-positive lists is to calculate a percentage of increase in efficiency of learning List 2 over List 1. The difference between both lists is first determined (positive or negative direction) and then the trials of List 1 are divided into this difference score. Due to the expected order effect this difference is likely to be positive, to a greater or lesser extent. Occasionally it is negative, as in the 13 instances cited above. However, after adding all changes algebraically it is possible to determine an overall improvement rate. When the order was positive to negative reinforcement value the nonspecific transfer amounted to 9%. However, when the order was negative to positive reinforcement value across lists the nonspecific transfer amounted to 38%.

DISCUSSION

The experimental hypotheses are taken as having received verification, with certain reservations as to the specific designation of overachievement and underachievement. Although the trends are in line with the designations as orginally conceived, it would appear that all low average to below average students are likely to respond with a rising positive reinforcement value facilitation effect. Since many of these low performers have the apparent ability—as judged by IQ tests—to do better, one does not find a collapsing of the positive reinforcement value effect with rising levels of intelligence. The reinforcement value effect is not tied to level of intelligence, at least in any obvious way. Furthermore, the fact that this effect can be shown across unmixed lists negates those criticisms which suggest that it is simply a mediational device which sorts out an opportunity for the subject to learn his "easiest" (liked) trigrams first. The conclusion is surely not difficult to draw that as a subject begins to fall off in level of performance—regardless of his potential—he reflects an increasing sensitivity to the affective aspects of a learning task.

In this regard, it is significant to note that O'Leary (1969) found inner-city Negroes responding more decidedly to reinforcement value affective factors than Caucasians of a comparable social class level. A negative affective attitude toward school-like tasks regardless of IQ factors may well be a significant dimension of the subject's performance. For example, the repressed black subject may be performing at a uniformly low level in school because of such attitudinal factors rather than limitations in native intellectual potential. This "emotionalized" side to human learning has not been given much attention as a basic explanatory variable in theories of learning, except in the sense that reinforcements have been thought of by the investigator as either positive or negative. If one compared the efforts which have gone into the study of transfer effects along such lines as the similarity of Lists A to B to C and so forth against the generalization of learning rates along a series of liked and/or disliked sequences there would surely be a decided contrast.

Efforts are underway today to develop learning proclivities in children who might be termed low performers, culturally deprived, or underachievers. The importance of transfer here along success versus failure lines may be crucial. For example, should one afford a deprived child 100% success experience, or might it be better to first determine his preferences, present him with a brief stint in an obviously difficult and hence failure-laden situation preliminary to a sequential and more extended experience with a preferred activity? The amount of benefit which might accrue to the latter tactic could well outweigh the 100% success (positive to positive transfer) condition. More work is needed on sequential effects of reinforcement value before an explicit delineation of such affective factors can be made. Work specifically done on repressed groups would be helpful, and of course at some point an extension must be made to actual school materials.

Finally, and more to the point of general research on the value of reinforcement, it seems evident that for some time psychologists have cultivated an unnatural, hence unprofitable, dichotomy between the laboratory and the applied or field setting. If it is possible to isolate a variable having relevance to the traditional interests of the field while meeting the strictures of the laboratory, then hopefully in time a more unified and relevant learning theory will be possible. Rather than adding to the shopworn dichotomies of "applied versus basic" or "laboratory versus personality" researches, the spirit of reinforcement value investigations has been to bring together all students of human behavior in a way which might well provide a common interest and ensure an ultimate achievement for all concerned.

REFERENCES

ABRAMSON, Y., TASTO, D. L., & RYCHLAK, J. F. Nomothetic vs. idiographic influences of association value and reinforcement value on learning. *Journal of Experimental Research in Personality,* 1969, **4,** 65–71.

ARCHER, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs,* 1960, **74**(10, Whole No. 497).

CIEUTAT, V. J., STOCKWELL, F. E., & NOBLE, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (*m, m'*) in paired associate learning. *Journal of Experimental Psychology,* 1958, **56,** 193–202.

COHLER, M. J. A comparative study of achievers and nonachievers of superior intelligence. *Summary of Doctoral Dissertations* (Northwestern University), 1940, **9,** 74–79.

CRONBACH, L. J. *Essentials of psychological testing.* New York: Harper, 1960.

DOWD, R. Underachieving students of high capacity. *Journal of Higher Education,* 1952, **23,** 327–330.

HARLOW, H. F. The formation of learning sets. *Psychological Review,* 1949, **56,** 51–65.

HOLLAND, J. L. The prediction of college grades from the California Psychological Inventory and the Scholastic Aptitude Test. *Journal of Educational Psychology,* 1959, **50,** 135–142.

HORRALL, B. M. Academic performance and personality adjustments of highly intelligent college students. *Genetic Psychology Monographs,* 1957, **55,** 3–38.

KAUSLER, D. H. *Readings in verbal learning.* New York: John Wiley & Sons, 1966.

LUM, M. A comparison of under- and overachieving female college students. *Journal of Educational Psychology,* 1960, **51,** 109–114.

MITCHELL, J. V. Goal setting behavior as a function of self-acceptance, over- and underachievement and related personality variables. *Journal of Educational Psychology,* 1959, **50,** 93–104.

MORGAN, H. H. A psychometric comparison of achieving and nonachieving college students of high ability. *Journal of Consulting Psychology,* 1952, **16,** 292–298.

O'LEARY, L. R. The effect of idiographic meaningfulness on the learning of subjects from different racial backgrounds, social classes, and ability levels. Unpublished doctoral dissertation, St. Louis University, 1969.

PEPPIN, B. H. Parental understanding, parental acceptance, and the self concept of children as a function of academic over and underachievement. *Dissertation Abstracts,* 1963, **23,** 4422–4423.

POSTMAN, L., & SCHWARTZ, M. Studies of learning to learn: I. Transfer as a function of method of practice and class of verbal materials. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3,** 37–49.

RINGNESS, T. A. Affective differences between successful and non-successful bright 9th grade boys. *Personnel and Guidance Journal*, 1965, **43,** 600–606.

RYCHLAK, J. F. Reinforcement value: A suggested idiographic intensity dimension of meaningfulness for the personality theorist. *Journal of Personality*, 1966, **34,** 311–355.

RYCHLAK, J. F. *A philosophy of science for personality theory*. Boston: Houghton Mifflin, 1968.

TENBRUNSEL, T. W., NISHBALL, E. R., & RYCHLAK, J. F. The idiographic relationship between association value and reinforcement value, and the nature of meaning. *Journal of Personality*, 1968, **36,** 126–137.

(Received January 29, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 2, 148–156

FEEDBACK PROCEDURES IN PROGRAMMED INSTRUCTION[1]

RICHARD C. ANDERSON,[2] RAYMOND W. KULHAVY, AND THOMAS ANDRE
University of Illinois

Three hundred fifty-six subjects in two experiments completed a programmed lesson on the diagnosis of myocardial infarction. The experiments were conducted using a computer-based instructional system which insured that, unlike many previous studies, the subject responded before he received knowledge of the correct response (KCR). The subjects who received KCR after they responded learned significantly more than subjects who received no KCR or who could peek at KCR before they responded. Neither two versions of a time-out procedure, a correction procedure, nor a forced-return procedure improved on KCR after responses. KCR after wrong answers only was slightly better than KCR after right answers only. Review of incorrect frames failed to improve performance on either an immediate or delayed criterion test.

One of the cardinal principles of programmed instruction is that immediate knowledge of the correct response (KCR) facilitates learning. Consequently, programmers were unpleasantly surprised when study after study showed that programs teach as well when the KCR after each frame is omitted (Feldhusen & Birt, 1962; Hough & Revsin, 1963; Krumboltz & Weisman, 1962; Moore & Smith, 1961, 1964; O'Day, Kulhavy, Anderson, & Malcynzski, 1971; Rosenstock, Moore, & Smith, 1965). Lublin (1965) found that a group that completed the first 1,052 frames of the Holland and Skinner (1961) program without KCR actually did significantly better on the criterion test than the group that saw the correct answer after every frame.

The usual assumption has been that KCR is reinforcing; however, there is good reason to believe that this assumption arises from an inappropriate analogy to the food- or water-deprived animal (cf. Anderson, 1967). Reinforcement maintains in greatest strength the response which immediately precedes it. If an explicit "observing response" (Holland, 1958) were required of students in order for them to receive KCR, the rate at which this observing response occurred would be a measure of the reinforcing value of KCR. Melching (1966) gave KCR only when requested to students completing a 364-frame program on magnetism. Feedback was requested on just 31% of the frames. Errors were made on 28% of the frames in which KCR was requested but on only 4% of the frames in which KCR was *not* requested. There are those who contend that when a student bypasses KCR he is engaging in "self-reinforcement" because he "knows he is right"; however, it can be argued that

> the only decent thing to do in the face of these data is to question the assumption the KCR is a reinforcer. After all, the food-deprived pigeon does not ignore the pellet of food even when it "knows food is available"—that is, when there is a reliable discriminative stimulus for the presence of food, such as the sound of the food magazine [Anderson, 1967, pp. 149–150].

If KCR is a reinforcer, then, by analogy with the animal literature, introducing a delay between the response and KCR should interfere with learning. No such interference appears. Indeed, a number of investigators have found that a delay between the response and KCR not only fails to impede learning but actually improves retention (e.g., Sassenrath & Yonge, 1969).

---

[1] The authors are grateful for the assistance of Gary Arsham, M.D., Sylvia Thornburg, Sheila Goldberg, and Maurice Mahan. The PLATO system is currently supported in part by the Advanced Research Projects Agency through the Office of Naval Research under Contract Nonr 3985(08), and in part by the National Science Foundation under Contract NSF GJ 81. The research described herein was completed under the former contract. A version of this paper was read at the annual meeting of the American Educational Research Association, Minneapolis, March 1970.

[2] Requests for reprints should be sent to Richard C. Anderson, Training Research Laboratory, University of Illinois, 8 Lincoln Hall, Urbana, Illinois 61801.

Several experiments employing self-instructional programs have presented KCR on intermittent schedules (Krumboltz & Weisman, 1962; Lublin, 1965; Rosenstock et al., 1965). The investigators were apparently aiming to produce the well-known effects of intermittent reinforcement, persistence, and resistance to extinction. The difficulty is that in programmed lessons the responses and the stimuli are continually changing and under these conditions intermittent reinforcement impedes learning. So, even if KCR is reinforcing, it is unclear what effect an intermittent schedule should have.

It is obvious that KCR can provide corrective feedback when the student makes a mistake. Buss and his associates (e.g., Buss, Braden, Orgel, & Buss, 1956) have found that either "right and "wrong" or no comment and "wrong" produced more rapid learning than saying "right" when the subject was correct and giving no comment when he was wrong. Travers, Van Wagenen, Haygood, and McCormick (1965) found no difference between saying "right" and giving no comment when the student was correct. When the student was incorrect, however, there was a big advantage to giving him the correct response in addition to telling him that he was wrong. These studies suggest that KCR functions more as corrective feedback than as reinforcement.

One of the purposes of the research described herein was to explore the function of KCR in programmed lessons. If KCR is chiefly important as reinforcement then the optimum procedure would be to differentially reinforce correct responses, that is, to present KCR after correct answers and withhold KCR after incorrect answers. On the other hand, if the main function of KCR is to provide corrective feedback, then presenting KCR after incorrect responses only should be as effective as furnishing KCR after every response.

Whatever the exact function of KCR, several lines of reasoning, as well as a great deal of empirical evidence from studies that have not involved programmed instruction, suggest that KCR should be facilitative. A plausible interpretation of the failure to find facilitation in programmed lessons is that students pay less attention when immediate KCR is provided. Sometimes there may be a short-circuiting of attention when the correct answer is readily available: the student may copy it into the blank without reading the material in the frame. At least two of the previous studies employed programmed texts in which the frames were printed in consecutive order down sheets of paper (Lublin, 1965; Rosenstock et al., 1965). The correct answers appeared immediately below the frames. This arrangement makes accidental exposure of the answers difficult to avoid, and of course makes deliberate cheating easy. It is known that students do not learn much simply from copying answers into blanks (cf. Anderson & Faust, 1967). Both of the experiments reported in this paper were completed using a computer-controlled system which made it possible to insure that KCR was not available until after the student responded. In addition, the second experiment directly investigated the hypothesis that many students turn a program into a sequence of "copying frames" when they have access to KCR before responding.

Elsewhere we have argued that students tend to follow the law of least effort, especially when they are tired, bored, or under pressure to work quickly (Anderson, 1970). In other words, it is argued that students will often fail to learn what a lesson is intended to teach them when it is possible to short-circuit the instructional task. Even if they are not actually copying from the KCR, students may study less carefully when KCR is always immediately available. There is some indirect evidence to support this proposition. Lublin (1965) found that the group that always received KCR completed the program in substantially less time than the other groups. Similarly, Sullivan, Baker, and Schutz (1967) found that a group which received KCR on tests inserted with a 60-page text on military justice performed no better on the criterion test than a group that did not receive KCR, but finished the text in less time. Sullivan et al.

(1967) interpreted their results to mean that

> "no feedback" Ss expended more time and effort attempting to learn from the prose textual material. For these individuals, of course, this [was] the only instructional material in the text. The "feedback" Ss, on the other hand apparently neglected the text material to some degree and used the instructional value of the immediate feedback...

If students become careless because of KCR, this could offset KCRs presumed beneficial effects. It is possible, therefore, that KCR would show to advantage if procedures were employed which maintained attention. Two such procedures were investigated in the present research. The first was a brief "time out" following errors. A delay is mildly aversive and may increase the incentive for careful attention to the material (cf. Zimmerman & Ferster, 1963). Based on the same reasoning, a correction procedure was also investigated. Under this procedure, the student had to respond correctly to each frame, making up to 10 responses if necessary, before he advanced to the next frame.

EXPERIMENT I

Method

Design and subjects. The experiment involved eight feedback procedures and three levels of verbal ability. The subjects were 168 summer students enrolled in an educational psychology course. Participation in the study was a course requirement. Prior to the experimental sessions all of the potential subjects were given the Educational Testing Service Wide Range Vocabulary Test (French, Ekstrom, & Price, 1963). On the basis of scores on this test, subjects were stratified into three levels of verbal ability. The subjects within each level were randomly assigned to feedback procedures. Ten subjects were replaced for failure to complete the program within 2 hours. Eighteen additional subjects participated in a pilot study.

Materials. The program was Tobias' (1968) modification of a linear program (originally prepared by Francis Mechner) on the diagnosis of myocardial infarction from electrocardiograms. The program was too long to complete within a 2-hour experimental session so 36 frames from a section on the incidence and causes of heart disease were removed. The form of the program used in the study entailed 112 frames averaging 22 words in length. Many frames also contained diagrams, ECG tracings, or tables. Each frame required the subject to respond to either a multiple-choice or completion question. In the version of the program used in this research students were not asked to construct drawings.

The criterion test contained 39 multiple-choice and completion items. Many of the items required multiple answers; hence, the total possible score was 84. Misspelled answers to completion items were counted right as long as the response was recognizable. None of the questions required the subject to construct drawings. Some of Tobias' (1968) items were used and the remainder were developed by the authors.

Feedback conditions. KCR was presented separately from either the preceding or following frame. In other words, a frame and KCR were never simultaneously exposed, though the KCR often repeated portions of the preceding frame. The feedback procedures were as follows:

0%-KCR: No KCR was presented after any frame.

100%-KCR: KCR was presented after every frame.

KCR-R: KCR was presented after right answers only.

10%-KCR-R: KCR was presented after a random 10% of correct responses; no KCR was given after incorrect responses.

KCR-W: KCR was presented after wrong answers only.

Time Out: There was a 15-second time out, with the frame in view, after wrong answers; then KCR was presented. No KCR was provided after right answers.

Correct: KCR was provided after each right answer. In the event of a wrong answer, the program looped back to the same frame with instructions to try again. The program advanced to KCR whenever a subject made 10 errors on the same frame.

Voluntary: After responding, the subject voluntarily selected KCR or no KCR. Pressing one button gave KCR; pressing the same button again led to the next frame. Pressing another button caused the next frame to appear immediately.

Procedure. The experiment was conducted on PLATO, a computer-based instructional system located at the University of Illinois (Bitzer, Hicks, Johnson, & Lyman, 1967). PLATO consists of a Control Data Corporation 1604 computer connected to 20 student stations housed in semienclosed cubicles. Each station contains a 12-inch square TV screen upon which stimuli are presented and a keyboard upon which responses are entered.

The subjects were run in groups of 10–18. Of course, PLATO allowed individualized treatment of each subject. All experimental groups were represented in most sessions. PLATO presented the instructions appropriate for each feedback condition, presented the program, and recorded errors during the program, time on each frame, and time on each KCR presentation. The time data were recorded in 60ths of a second and were probably accurate to one-third of a second. When a subject completed the program he was directed to a nearby room where the criterion test was administered in paper-and-pencil form.

Results

Table 1 contains mean correct on the criterion test, mean program errors, mean time per frame, and mean time per KCR presentation. There were significant differences on the criterion test among feedback procedures, $F = 3.37$, $df = 7/144$, $p < .01$, and among verbal ability levels, $F = 3.76$, $df = 2/144$, $p < .05$. The interaction between these factors was not significant. Newman-Keuls tests showed that the 100%-KCR, Time out, Correct, and Voluntary groups all did significantly better ($\alpha = .05$) on the criterion test than the 0%-KCR group. No other comparisons were significant.

The differences among feedback procedures in mean frequency of program errors (counting only the first error to any frame in the Correct group) were not significant, $F = 1.93$, $df = 7/144$. Verbal ability was related to program errors, $F = 11.82$, $df = 2/144$, $p < .01$. There was also a Feedback Procedure × Ability interaction, $F = 2.16$, $df = 14/144$, $p < .05$, primarily because of the relatively high frequency of errors on the part of low-ability subjects in groups 0%-KCR, 100%-KCR, and KCR-R.

Program errors were considerably more frequent than expected, due in part to the fact that PLATO was programmed to count typographical errors and spelling mistakes. This fact may have undermined the performance of the Time Out and Correct groups since delays or repetition because of trivial mistakes undoubtedly would be annoying. Incidentally, the Correct and Time Out groups spelled correctly 94% of the technical terms on the criterion test whereas the average for the other groups was 85%.

There were no significant differences, $F = 1.98$, $df = 7/144$, among the feedback procedures in mean time per frame (including only the first exposure times in the Correct group). Time per frame was a decreasing function of verbal ability, $F = 6.11$, $df = 2/144$, $p < .01$. There were significant differences, $F = 13.79$, $df = 6/126$, $p < .01$, among the feedback procedures in mean time per KCR exposure, excluding the 15-second time out in the Time Out group. Newman-Keuls tests showed that the Time Out group spent significantly more time ($\alpha = .05$) per KCR than all groups except the KCR-W and the Voluntary groups. The KCR-W and Voluntary groups spent more time than KCR-R, Correct, and 100%-KCR. Group KCR-W also spent more time than 10%-KCR-R. No other comparisons were significant.

EXPERIMENT II

The chief purpose of Experiment II was to replicate, if possible, the finding of Experiment I that students who get KCR learn more than students who do not, and to test the hypothesis that KCR is facilitative only when students cannot peek at it before responding. Also investigated were the effects of delayed review of frames upon which errors were made, and two procedures intended to maintain attention by punishing students for errors.

TABLE 1

MEANS FOR EACH FEEDBACK PROCEDURE IN EXPERIMENT I

Measure	0%-KCR	100%-KCR	KCR-R	10%-KCR-R	KCR-W	Time Out	Correct	Voluntary
Criterion test correct	57.9	68.7	61.8	61.3	64.5	67.0	68.5	66.2
Program errors	32.8	29.0	34.8	34.7	31.6	25.7	26.0[b]	32.2
Time per frame (sec.)	31.2	29.7	28.6	30.9	29.1	30.4[a]	23.9[b]	28.1
Time per KCR (sec.)	—	4.0	3.3	5.8	8.7	10.2[a]	3.9	7.5

Note.—There were 21 subjects in each group, 7 at each ability level.

[a] Does not include 15-second penalty for wrong answers.

[b] Includes only the first exposure to any frame.

Method

Design and subjects. The experiment entailed an uninstructed control group, six feedback procedures, and three levels of verbal ability. The subjects, who were randomly assigned to procedures from within ability levels, were 188 students enrolled in an introductory educational psychology course. Two additional subjects were dropped for failure to complete the program within 2 hours and the 19 subjects in one session were lost when the treatment was not implemented properly because of modifications which had been made in the systems program.

Materials. Based in part on the data from Experiment I, a physician and an experienced programmer modified both the program and the criterion test. Nineteen frames in the program were consolidated or deleted. Eighteen frames were revised slightly. Five new frames were added. The version of the myocardial infarction program used in Experiment II consisted of 98 frames averaging 23 words in length and requiring a total of 126 separate responses. Forty-five frames included drawings as well as text.

The criterion test was modified on the basis of an item analysis and judgments about validity. Twelve items were deleted and nine new items constructed by the authors were added. The final test contained 36 multiple-choice and completion questions yielding a possible score of 66.

Feedback conditions. KCR appeared in the lower right-hand corner of each frame below three rows of Xs, rather than on a separate slide as in the initial study. Both the frame text and the KCR were in view when KCR was presented. The feedback procedures are described below.

0%-KCR: Same as Experiment I.

100%-KCR: Same as Experiment I.

Forced: KCR was presented immediately after the response to each frame. Following KCR on error frames, the subject was forced to repeat the frame until he responded correctly.

Review: KCR was presented immediately after the response to each frame. Frames upon which errors were made were repeated after sections 28, 24, 19, and 27 frames in length. Frames still in error were then presented again. Review continued until each frame had been answered correctly, at which point the next section began.

Time Out$_2$: KCR was presented after every frame. There was a 15-second time out, with the frame text and KCR in view, whenever an error occurred.

Peek: KCR was presented at the same time the frame was exposed. In other words, the subject had access to KCR before he responded. Initial instructions repeatedly stressed that the subject should respond before he looked at the KCR.

Control: This was an uninstructed control which received only the criterion test.

Procedure. In order to reduce the number of unimportant program errors, PLATO was programmed to count common misspellings of technical terms as acceptable responses. To obtain an estimate of retention, the criterion test was readministered 14–16 weeks after the experimental sessions to those subjects still available. The procedure of the two experiments was identical in all other respects.

Results

Table 2 contains mean correct on the criterion test, mean program errors, mean time per frame, and mean time per KCR presentation for the groups in Experiment II. An unweighted means analysis of criterion test variance, which did not include the uninstructed control group, revealed significant differences among feedback procedures, $F = 3.30$, $df = 5/128$, $p < .01$, and ability levels, $F = 4.99$, $df = 2/128$, $p < .01$. As predicted, the 100%-KCR group learned more than either the 0%-KCR, $t\ (48) = 2.16$, $p < .05$, or Peek group, $t\ (44) = 3.83$, $p < .01$. Newman-Keuls tests showed no further significant ($\alpha = .05$) differences.

Analysis of program errors yielded significant effects for both feedback proce-

TABLE 2

MEANS FOR EACH FEEDBACK PROCEDURE IN EXPERIMENT II

Measure	0%-KCR	100%-KCR	Forced	Review	Time Out$_2$	Peek	Control
Number of subjects	26	24	25	25	24	22	42
Criterion test correct	47.4	52.8	49.6	47.5	49.9	43.3	16.4
Program errors	31.2	27.7	28.5[a]	27.9[a]	26.5	14.9	—
Time per frame (sec.)	28.7	24.7	[b]	22.8[a]	23.5[c]	23.9	—
Time per KCR (sec.)	—	3.2	[b]	2.7[a]	3.3[c]	—	—

[a] Includes only the first encounter with any frame.

[b] No data collected.

[c] Does not include 15-second penalty.

dures, $F = 9.18$, $df = 5/128$, $p < .01$, and ability level, $F = 14.19$, $df = 2/128$, $p < .01$; however, unlike the first experiment, the interaction of these two factors was not significant. Newman-Keuls tests showed that the frequency of program errors was significantly lower ($\alpha = .05$) in the Peek condition than in any other. None of the remaining comparisons was significant.

Ignoring Forced subjects, on whom data were not collected, and including only the first exposure of frames in the Review group, analysis of time per frame showed significant effects for feedback procedure, $F = 5.15$, $df = 4/106$, $p < .01$, and verbal ability, $F = 5.33$, $df = 2/106$, $p < .01$. Newman-Keuls tests showed that the 0%-KCR group spent significantly more time per frame than the other groups. No other comparisons were significant.

An analysis of time per KCR exposure showed no significant effect due to feedback procedure or ability.

There were no differences among feedback procedures on the delayed criterion test; however, all the groups that had received the program performed significantly better than the uninstructed control group.

DISCUSSION

The most important result of the experiments described in this paper was the findings that students who received KCR after every frame performed better on the criterion test than students who received no KCR. To our knowledge, only one other experiment has ever found that KCR facilitates learning from a programmed lesson (Meyer, 1960). In many of the previous studies it was possible for the student to copy the correct answer into the blank without studying the frame. This, we believe, was a major reason for poor results using KCR. The present experiments were completed on a computer-controlled instructional system which made it possible to insure that the student composed a response before he saw KCR.

The hypothesis is that many students will turn a program into a series of "copying frames" when it is possible for them to see KCR before responding. This hypothesis was directly investigated in Experiment II. Despite the fact that the instructions for the Peek condition stated three times that the student should enter his answer before he looked at the correct answer, the Peek group learned substantially less than the 100%-KCR group, indeed, even less than the 0%-KCR group. The Peek group averaged about half as many program errors as the other groups. Merely receiving KCR did not reduce errors since errors were almost as frequent in the 100%-KCR group as they were in the 0%-KCR group. Therefore, the very low frequency of errors in the Peek group must have been due to copying. After the experiment students were asked to make any comment they wished about the program. Eleven of the 22 subjects in the Peek group mentioned looking at the KCR before responding. Sample comments: (*a*) "It's hard not to look at the answers. It would be better if they were given in the next frame." (*b*) "The answers should not be shown immediately, as this causes cheating to be easier." (*c*) "I enjoyed it. One problem, however, I could not help seeing some of the answers out of the corner of my eye, even though I tried not to look." (*d*) "Very interesting, but I tend to look at the answers too soon."

It may be questioned whether the results obtained under the Peek condition are generalizable to program formats in which KCR is less prominently displayed. Programmed texts printed in the "down the page" format are probably quite susceptible to copying. Even when a mask is provided to cover the answers, the student may not use the mask, he may inadvertently expose the KCR too soon, or, of course, he may cheat. The student must turn the page to find KCR when studying a programmed text arranged in the "across the page" format. This arrangement precludes accidental exposure of KCR but it does not prevent cheating. Premature exposure of KCR is probably not uncommon in programmed texts no matter what the format. Anderson (1969) evaluated the effectiveness of a programmed text on population genetics in 15 high school biology classrooms. To avoid accidental exposure of KCR, each frame occupied a separate page; the KCR for that frame appeared at the top of the next page. For the same reason, the program was mimeographed on blue paper through which the KCR could not be read (cf. Anderson, Faust, & Roderick, 1968; Brown, 1966). When they completed the program students were asked how frequently they had turned the page and copied the correct answer when they found a question difficult. Despite the fact that the directions to the program exhorted students to "*write* an answer to each question and fill in each blank *before* you look ahead to the correct answer," better than 40% of the students said they sometimes copied answers to difficult questions and 20% said they usually did. Achievement was a decreasing function of reported frequency of copying.

Insuring that the student responds before KCR is provided will prevent him from turning a program into a series of copying frames, but such control will not help if the program already consists largely of copying frames. Figure 1 contains three sample frames from a program on how to program. It could not possibly make any difference whether KCR were furnished after either of the first two frames, because the frames themselves contain the answers. KCR would also be irrelevant after the third frame if it appeared immediately after the other two frames, since these frames comprise a strong "sequence prompt." Several studies using teaching machines have failed to show a significant advantage for KCR, despite the fact that the machine presumably prevented the student from seeing KCR before he responded (Feldhusen & Birt, 1962; Hough & Revsin, 1963; Moore & Smith, 1961, 1964). Our best guess is that KCR failed to show to advantage because the programs used in these studies contained many copying frames and were otherwise heavily prompted.

One of the purposes of Experiment I was to investigate whether KCR functions as reinforcement or corrective feedback. For what it is worth the KCR-W group performed somewhat better on the criterion test than the KCR-R group; however, the difference was not significant. Neither the

```
S-6.  There are a number of assumptions we (the programmers)
      must make about you (the learner).  One is that you
      read at a certain level and have a vocabulary consistent
      with this program.  Programmers must make ____________
      about the learner.

R-6.  assumptions
```

```
S-8.  We, as programmers, not only have to make certain assump-
      tions about you, the learner, but we also must determine
      the goals or objectives of the program.  The goals or
      ____________ as well as the assumptions are integral
      parts of programming for automated instruction.

R-8.  objectives
```

```
S-13. Two important considerations of "programmed learning" are
      to make (a)____________ about the learner and to deter-
      mine the (b) ____________ of the program.

R-13. (a) assumptions
      (b) objectives
```

FIG. 1. Three frames from a program on how to program. (Reprinted with permission from *A Guide to Programmed Instruction* by J. P. Lysaught and C. M. Williams. Copyrighted by John Wiley & Sons, Inc., 1963.)

KCR-W nor KCR-R group was significantly different from the 0%-KCR or 100%-KCR group.

Two kinds of time out procedure, a correction procedure, and a forced return procedure were investigated under the assumption that mild punishment for errors would cause subjects to be more careful. None of these procedures improved upon 100%-KCR.

The Voluntary condition was included in Experiment I in order to try to replicate Melching's (1966) finding that subjects choose KCR on only a small proportion of frames. Subjects in the present study selected KCR on 95% of the frames. To advance to the next frame, a subject could either press the KCR button twice in rapid succession or press an alternative button once to advance directly. It is suspected that most subjects used the former strategy, which would explain the difference between the present results and Melching's.

Holland and Porter (1961) reported that delayed review of frames upon which errors were made improved retention when the criterion test was readministered after 6 months. The present research failed to replicate this finding. The Review group performed slightly worse than the 100%-KCR group on both the immediate and delayed criterion test.

REFERENCES

Anderson, R. C. Educational psychology. *Annual Review of Psychology*, 1967, **18**, 129–164.

Anderson, R. C. The comparative field experiment: An illustration from high school biology. *Proceedings of the Invitational Conference on Testing Problems*. Princeton, N.J.: Educational Testing Service. 1969.

Anderson, R. C. Control of student mediating processes during verbal learning and instruction. *Review of Educational Research*, 1970, **40**, 349–370.

Anderson, R. C., & Faust, G. W. The effects of strong formal prompts in programed instruction. *American Educational Research Journal*, 1967, **4**, 345–352.

Anderson, R. C., Faust, G. W., & Roderick, M. C. "Overprompting" in programmed instruction. *Journal of Educational Psychology*, 1968, **59**, 88–93.

Bitzer, D. L., Hicks, B. L., Johnson, R. L., & Lyman, E. R. The PLATO system: Current research and developments. *IEEE Transactions on Human Factors in Electronics*, 1967, **8**, 64–70.

Brown, R. W. Format location of programed instruction confirmations. *Journal of Programed Instruction*, 1966, **3**, 1–4.

Buss, A. H., Braden, W., Orgel, A., & Buss, E. Acquisition and extinction with different verbal reinforcement combinations. *Journal of Experimental Psychology*, 1956, **52**, 288–295.

Feldhusen, J. F., & Birt, A. A study of nine methods of presentation of programmed learning material. *Journal of Educational Research*, 1962, **55**, 461–466.

French, J. W., Ekstrom, R. B., & Price, L. A. *Kit of Reference Tests for Cognitive Factors*. Princeton, N. J.: Educational Testing Service, 1963.

Holland, J. G. Human vigilance. *Science*, 1958, **128**, 61–67.

Holland, J. G., & Porter, D. The influence of repetition of incorrectly answered items in a teaching machine program. *Journal of the Experimental Analysis of Behavior*, 1961, **4**, 305–307.

Holland, J. G., & Skinner, B. F. *The analysis of behavior*. New York: McGraw-Hill, 1961.

Hough, J. B., & Revsin, B. Programmed instruction at the college level: A study of several factors influencing learning. *Phi Delta Kappan*, 1963, **44**, 286–291.

Krumboltz, J. D., & Weisman, R. G. The effect of intermittent reinforcement in programmed instruction. *Journal of Educational Psychology*, 1962, **53**, 250–253.

Lublin, S. C. Reinforcement schedules, scholastic aptitude, autonomy need, and achievement in a programmed instruction course. *Journal of Educational Psychology*, 1965, **56**, 295–302.

Lysaught, J. P., & Williams, C. M. *A guide to programmed instruction*. New York: Wiley, 1963.

Melching, W. H. Programed instruction under a feedback schedule. *National Society for Programed Instruction Journal*, 1966, **5**, 14–15.

Meyer, S. R. Report on the initial test of a junior high school vocabulary program. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning*. Washington, D. C.: National Education Association, 1960.

Moore, J. W., & Smith, W. I. Knowledge of results in self-teaching spelling. *Psychological Reports*, 1961, **9**, 717–726.

Moore, J. W., & Smith, W. I. Role of knowledge of results in programmed instruction. *Psychological Reports*, 1964, **14**, 407–423.

O'Day, E. F., Kulhavy, R. W., Anderson, J. W., & Malczynzski, R. *Programmed instruction: Techniques and trends*. New York: Appleton-Century-Crofts, 1971, in press.

Rosenstock, E. H., Moore, J. W., & Smith, W. I. Effects of several schedules of knowledge of results on mathematical achievement. *Psychological Reports*, 1965, **17**, 535–541.

Sassenrath, J. M., & Yonge, G. D. Effects of delayed information feedback and feedback cues in learning on delayed retention. *Journal of Educational Psychology*, 1969, **60**, 174–177.

Sullivan, H. J., Baker, R. L., & Schutz, R. E.

Effects of intrinsic and extrinsic reinforcement contingencies on learner performance. *Journal of Educational Psychology*, 1967, **58,** 165–169.

Tobias, S. *The effect of creativity, response mode, and subject matter familiarity on achievement from programmed instruction.* New York: MSS Educational Publishing Company, 1968.

Travers, R. M. W., Van Wagenen, R. K., Haygood, D. H., & McCormick, M. Learning as a consequence of the learner's task involvement under different conditions of feedback. *Journal of Educational Psychology*, 1964, **55,** 167–173.

Zimmerman, J., & Ferster, C. B. Intermittent punishment of S^{Δ} responding in matching to sample. *Journal of the Experimental Analysis of Behavior*, 1963, **6,** 349–356.

(Received February 17, 1970)

Journal of Educational Psychology
1971, Vol 62, No. 2, 157-166

SELECTION AND SOCIALIZATION:

A STUDY OF ENGINEERING STUDENT ATTRITION[1]

ROBERT ATHANASIOU[2]
Johns Hopkins University

An 18-month cross-validated panel study of virtually an entire class of engineering students indicated that change scores on indexes of professional and educational interests, self-perception, and authoritarianism were better predictors of attrition than either initial or final index scores. Attrition from engineering was defined by transfer to another field of study within the university. Changes on each index followed the predicted pattern of accentuation with high scores getting higher and low scores getting lower. The small differences, at entrance, between those who stayed in engineering and those who transferred, and the rather large differences at exit suggest that early identification (or selection) programs might not be as effective as changes in the socializing environment in reducing engineering student attrition.

Selection and socialization are two rather broad terms used in this study. Selection is defined as the process wherein an institution recruits and admits those individuals whose characteristics at entrance are presumed to represent a good fit with organizational requirements. The degree of success an institution achieves in recruiting those individuals who will eventually meet its criteria (e.g., graduation, in this case) is the degree of predictive validity attributed to the selection process.

The socialization process, as used in this study, refers to the experiences, rewards, etc., which an individual encounters within an institution which serve to direct him toward a set of values, behaviors, and skills which are congruent with the institution's purpose. The effect of socialization relative to selection may be inferred by the difference between predictive and concurrent validity. Concurrent validity refers to the assessment of present status.

Authors such as White (1952), Keniston (1965), and Sanford (1962, 1966, 1967) have given attention to the question of individual personality change in late adolescence. Personality change in late adolescence opens the door to the general question of adult socialization. Brim and Wheeler (1966) deal with the need for adult socialization associated with progress through the life cycle. The emphasis on socialization seems to be a relative newcomer to the study of college students (however, see Newcomb, 1943).

The American College (Sanford, 1962) represents one of the first large-scale attempts to integrate social and personality theory with the study of higher education. In this massive work one notes a number of chapters which emphasize the process of selection (see Ch. 17, 19, 20). Additionally, articles by Beal and Bordin (1964), Kulberg and Owens (1960), and Mosier and Kuder (1949) also deal with selection as a major factor in successful vocational choice. This emphasis on selection gives way, however, when one considers *Where Colleges Fail* (Sanford, 1967), to a pronounced emphasis on socialization. The failure of colleges lies in their inability to adequately change students in areas such as social responsibility, academic achievement, creativity, etc. Kanter (1967) has suggested that socialization may prepare the student for participation in the appropriate adult occupational role.

---

[1] The research reported in this paper stems in part from a doctoral dissertation submitted by the author to the Doctoral Program in Social Psychology, Horace H. Rackham School of Graduate Studies, University of Michigan, Ann Arbor. The study was funded by the office of Manpower Administration, United States Department of Labor under Doctoral Dissertation Grant No. 91-24-68-61. The author wishes to acknowledge his debts to the late Stuart A. Kanter who directed the early data collection phase of this study and to his dissertation chairman, Theodore M. Newcomb, who provided many helpful suggestions throughout the course of this study.

[2] Requests for reprints should be sent to Robert Athanasiou, Department of Psychology, the Johns Hopkins University, Baltimore, Maryland 21218.

Feldman and Newcomb (1969) see the content of the selection and socialization processes as varying between colleges and between major fields.

In this last viewpoint there is somewhat greater perspective and differentiation than in the others. The recognition of initial differences (selection differences) and their effect on subsequent differences due, presumably, to socialization, adds considerable breadth to the subject of adult socialization. By examining differences between institutions or college majors one may be able to see the differential effects of role demands, organism capacity, social environment, etc., on the outcome of the socialization process.

The most recent and undoubtedly the most comprehensive evaluation of studies dealing with individual changes occurring during the college years is *The Impacts of Colleges upon Their Students* by Feldman and Newcomb (1969). In their analysis of the findings of more than 1,300 studies there is one process which seems to show up with great regularity. They comment,

> What we discovered to be most usual, in the studies we have surveyed, is that diversities among entering student bodies are, if anything, amplified during the college years—a phenomenon not explainable in terms of the artifacts of regression and related effects [p. 142].

This amplification of diversities usually takes the form of increasing scores on the variable of interest on the part of those subjects who were already high scorers. This phenomenon is called accentuation by Feldman and Newcomb and refers to "... an initially favorable or unfavorable attitude that becomes more so over time [p. 143]."

In general, their discussion is devoted to a consideration of differences between major field or colleges. Their assessment of the etiology of this accentuated diversification is that it is primarily a function of environmental differences. Their examples of the phenomenon:

> ... reveal no general tendency for individuals, independently of colleges attended, to increase their initial diversities. Rather, as in the case of academic majoring ... it is a matter of "fit" between the individual and the institutionally provided environment—a fit that represents a reciprocal adaptation of differential selection of student recruits and of environmental characteristics provided by the institution [p. 142].

In other words, socialization is thought to proceed on the basis of selection characteristics and the character of socialization is determined by environmental exigencies. This position is quite useful when the differences between colleges or majors, that is, the differences between environments, are to be placed in their appropriate perspective. It is not useful if there is a need to explain the accentuation of differences within a single environment or academic major.

Kanter (1967) has clearly shown that there are rather large differences between those who stay in a major field (specifically, engineering) and those who elect to leave if after three semesters. However, the question of selection differences was left unanswered and the basic assumption presented by Feldman and Newcomb was not confronted. Given the Feldman-Newcomb paradigm, one could only hypothesize that those students in the Kanter study who left the field (engineering) had simply made an inappropriate vocational choice and that their differences at exit were essentially the same as those at entrance. Such an error in vocational choice would then be readily attributable to errors in selection resulting in a poor "fit" between the individual and the environment. That is, students who transfer are "atypical" and presumably were so when they entered. Data presented here tend not to confirm this hypothesis. How, then, might differences among students arise?

Swap (1968) has considered the process of within-group accentuation to be mediated by reference groups and predicts that those values which will be accentuated are those which are important elements in the socialization process within the group. Moreover, those values which are complemented by the role requirements for being a member of the group are likely to be accentuated. Engineers, then, could be expected to adapt to role requirements such as impersonality and preference for structure, to lean toward intolerance of ambiguity, preference for seeking practical solutions, etc. The faculty, the environment, and the academic requirements of the engineering school, as detailed by Kanter (1967, pp. 29–35), would all provide congruent pressures in these directions. This does not, however, explain or even suggest why students who eventually transfer might show changes in the opposite direction.

Athanasiou (1968b) has briefly discussed changes shown by transfers on an empirically derived scale consisting of Omnibus Personality Inventory items. These findings indicate a significant degree of accentuation for transfer students as well as small but significant initial differences between transfers and engineers. Throughout the remainder of this paper the term "transfer" will refer to students who left the engineering school for another college in the university after at least one semester. "Engineers" will refer to students who stayed in the engineering school for at least three semesters and indicated that they were at least somewhat satisfied with engineering.

Here, then, is some indication of accentuation of differences within a major field. Moreover, it appears that these students were not very "atypical" at entrance, although there were such entrance variables as religion, sibship, and verbal aptitude which did discriminate between engineers and transfers.

The picture of accentuation which begins to take form is colored by some seemingly discrepant findings. As cited above, there is evidence for accentuation of differences both within and between majors. Environmental differences and selection seem adequate to explain differences between majors. Selection, if it were operating, might explain some of the within-major accentuation, but it seems to have a minimal effect. In engineeering, the curricular environment is essentially the same for transfers and engineers. The environment, however, consists of more than the curriculum and there may be some small differences between the groups at entrance. These differences may be accentuated if they involve values, attitudes, etc., which are important to the individual or are congruent with situational role demands.

In college, students contact the socializing environment "... through interaction with others who hold normative beliefs about what his role should be and who reward and punish him for correct or incorrect actions [Brim, 1966, p. 9]. If the "others" (e.g., upperclassmen, faculty, etc.) with whom a potential transfer interacts are different from those with whom the engineer interacts, one may reasonably expect the content of the socialization process to differ.

Socialization can also take place through participation in organized activities. The expectations for appropriate behavior in this case are derived from the social system rather than from individuals.

METHOD

Virtually the entire Engineering College class of 1969 of a large midwestern university participated in the study just prior to beginning their freshman year in the fall of 1965. Of the 892 students in this group to whom questionnaires were mailed, 149 failed to respond and some 30 respondents were excluded on a priori basis due to their sex, non-United States residence, or age at time of entrance. The data from the remaining 713 subjects consisted of responses to a comprehensive questionnaire and the Omnibus Personality Inventory (OPI).

In the second semester of this group's sophomore year (winter 1967) a second, "sophomore," questionnaire was administered to all of the students presently at the university. At this time the effective population was 667 subjects, since 195 of the original group had left the university. An engineering version of the sophomore questionnaire was sent to those students still in the engineering college; students who had transferred out of the college received a slightly different transfer version. The lowest subgroup (sophomore transfers) return rate for questionnaires was 79.1%.

At the time of the second questionnaire administrations, 98% of the subjects were between the ages of 19 and 21. The modal residential area of their families was suburban (29%), with an additional 8% from farm families, and an additional 17% from large metropolitan areas. Seventy-one per cent of the sample had fathers whose occupational category could be rated middle-class or higher. The fathers showed a high degree of education, with 65% having some college or more.

The subjects reported that 73% of them had been in the top 10% of their high school class. Protestants comprised 63% of the sample, Catholics 20%, Jews 8%, and atheists or agnostics 9%. A majority (54%) were firstborn children.

Questionnaire Items and Hypotheses

Contained in the questionnaires[3] and other data were a large number of items which had been administered in each wave of testing. These items could be easily grouped into four content-quality categories:

1. The first subset of 122 items are those drawn from the OPI and are attitude-personality items which readily lent themselves to use as an authoritarianism scale. Items in this group were, for example, "Facts appeal to me more than ideas (true); I like modern art (false)." Forty-eight of the original 122 items remained after item analysis procedures described below.

2. Under the second rubric are 19 questionnaire items which deal with the individual's perception of himself and his personal goals. One item from this group was, "How much do you feel that the type of work you have chosen expresses your particular talents and interests?" Seventeen items were retained for this index.

3. The third category contains 41 items dealing with the individual's relationship to others and his views of friendship. A sample item from the friendship group was, "To what extent do you feel that a person should try to become close friends with others?" This index used 25 items from the original pool of 41.

4. The fourth category of 27 items is concerned with professional and educational interests. Items in this group asked the subject to compare engineering with other professions and also inquired about the subject's interest in courses outside the engineering school. Twenty-four of these items were retained for index building after item analysis. Hypotheses relative to each category of items are discussed below.

The relevance of the construct of authoritarianism to the present study lies in those segments of the syndrome which deal with intolerance of ambiguity, categorical thinking, responsiveness to the demands of authority, and anti-intraception. The education of engineering students, as currently constituted in many colleges and universities, seems both in structure and content to favor authoritarian modes of thinking and behavior. For example, the curriculum is often thoroughly programmed and highly structured; the time allocated for courses in the humanities is severely limited; the course material in engineering is nonambiguous in the sense that nearly every problem has a single correct answer, discoverable through the application of appropriate formulae; the subject matter, and often the instruction, are highly impersonal and nonsubjective; and there is little time in the engineering student's schedule for involvement with people or ideas. It seems reasonable to predict, therefore, that the orientation of an authoritarian individual should mesh rather well with the requirements of engineering training.

Such a prediction is somewhat at odds with the usual observation that college students are rather liberal and tend to become more so as their college career progresses. Feldman and Newcomb (1969) have extensively documented this general trend toward more liberal (pp. 338–443) less authoritarian (pp. 368–375) attitudes, and show the trend to be strikingly consistent and nearly universal. Somewhat to their surprise, however, they found no studies of change in authoritarianism among engineering students. Does the engineering curriculum recruit students high in authoritarianism, or does it make them that way after they have entered? Other studies by Athanasiou (1968a, 1968b) have hinted that attitudinal differences between engineers and potential transfers are significant, but quite small at entrance and increase through apparent changes in the transferring students' attitudes, while engineers' scores remain relatively stable. This would seem to suggest that the potential transfers become less authoritarian and leave engineering, while engineers find the environment reasonably supportive.

Items in the index of self-perception and perception of goals seem quite relevant to the supposedly developmental character of a college education. Beal and Bordin (1964), however, have indicated that the engineering student is more concerned with adjusting rather than growing. His efforts are devoted to satisfying "...mainly impersonal curiosities." His goals, and the rules for achieving them, are clear-cut, practical, thing-oriented rather than people-oriented, and in line with traditional masculine character. Kanter (1967) has shown that striking differences exist between engineers and transfers in their sophomore year with regard to the degree of introspection, self-analysis, and self-expression shown by his subjects. One may ask, however, whether these differences existed at entrance or were developed after contact with the engineering school environment.

Items dealing with interpersonal social relationships and views of friendship comprised the third index. In line with the impersonal, anti-intraceptive, "masculine" orientation of engineers as described in the two preceding statements of hypotheses, it is reasonable to expect that transfers will come to place a greater value on friendship per se than engineers. Furthermore, it is likely that they (transfers) will value in friendship the more intangible qualities like emotional support, closeness, warmth, openness, etc., which run counter to the rather low affective character of engineers (cf. Kanter, 1967).

---

[3] The questionnaires, coding directions for the indexes, and other data may be obtained by ordering a copy of "Selection and Socialization." Order No. 69-17, 959 from University Microfilms, Ann Arbor, Michigan, 48104. Microfilm cost is $3.00 and Xerox, $9.45.

The fourth category of items is more obviously related to the transfer–engineer distinction than the two previous categories. The "obtrusiveness" of these items (e.g., "...how would you compare the engineering profession with [other occupations]?") may operate to introduce a bias in the sophomore responses. As a consequence of possible dissonance effects (Brehm & Cohen, 1964), this index may be expected to show rather large differences at exit. On the other hand, it is unlikely that substantial differences will appear at entrance. Presumably, it would be discrepant for an entering freshman to hold a low opinion of engineering—the subject toward which he was about to devote 4 years of study. Moreover, virtually none of the subjects indicated that they had entered engineering with the intent of transferring at a later date.

Given these four indexes, the following hypotheses may be stated:

1. At entrance there will be relatively little difference, or no difference, between engineers and transfers; transfers may show slightly lower scores on the authoritarian index and slightly higher scores than engineers on all other indexes.

2. Engineers' scores will be more stable over the 18-month period than transfers' and the between group differences in index scores will account for a greater proportion of criterion variable variance at exit than at entrance.

3. Change scores in the direction of lower scores on the authoritarian index and increases on the other indexes will characterize transfers as opposed to engineers. Moreover, change scores will account for a greater proportion of criterion variable variance than either entrance or exit scores.

To this point the focus of this article has been directed toward setting the stage for a demonstration of the difference between predictive (selection) and concurrent (socialization) validities in a paradigm which assumed that accentuation takes place. While the loci and directions of accentuation have been discussed, the etiology of these expected changes should also be examined.

The sophomore and prefreshman versions of the questionnaire contain items which relate to a variety of potential socializing agents. Those items in both the sophomore and prefreshman questionnaires which relate to interaction with peers and authority figures who could reasonably be expected to hold normative expectations relevant to transferring behavior, items dealing with participation in organized activities, and items which deal with the individual's reaction to this educational environment (e.g., satisfaction with freshman counseling) were combined to form an index of socialization. Examples of items in this group were, "What percentage of your friends ... are also in engineering?" and, "Are you a member or an active participant in any engineering organization or activity?"

In line with statements made above, it is further hypothesized that:

4. The index of socialization should share a substantial proportion of common variance with change scores on the authoritarian, friendship, self-perception, and education indexes. Additionally, the socialization index itself will share some common variance with the criterion variable.

Index Construction

The basic method used to construct and test the previously described indexes was cross-validated item analysis (see Guilford, 1954).

The scoring formulae for the indexes were determined by examining only the prefreshman responses of a random sample (i.e., the analysis group) of subjects vis-à-vis the criterion variable of attrition. Each questionnaire item was examined for its ability to predict (at $p < .10$) attrition and scored accordingly using a dichotomous code. That is, if a "yes" response to a particular item was characteristic of transfers then "yes" was coded "1" while "no" and "maybe" were coded zero. The indexes were then scored for a separate holdout group of subjects by counting the frequency of the appropriate dichotomous item codes, as suggested by Bereiter (1967). A correction for guessing was applied to all index scores. These dichotomous item codes reduced all data to nominal characteristics and facilitated the computation of change scores. As noted previously a substantial proportion of items from each of the separate item pools was retained for index construction.

The index of authoritarianism was constructed in a slightly different manner. OPI item responses of the analysis group were compared with a subject's position in the upper and lower thirds of a measure of authoritarianism administered in the sophomore questionnaire (see Athanasiou, 1968b). Those OPI items which yielded a significant chi-square and an index of predictive association, λ, greater than or equal to a 15% reduction in error were retained, scored dichotomously, and combined to form an index of authoritarianism.

Change scores on the indexes were calculated for each subject at the time when the indexes were scored for the holdout group. An individual's change score on an index was his final (sophomore) index score minus his initial (prefreshman) index score (cf. Bereiter, 1967, pp. 10–11).

This procedure represents a somewhat unorthodox approach in that psychometric methods of test construction were applied to survey data. These methods tend to yield rather conservative results and may result in an underestimate of validity (see Dawes & Meehl, 1966).

TABLE 1

MEANS OF INDEXES BY CRITERION GROUPS

Group	Index and testing wave							
	Friendship		Self-perception		Education		Authoritarianism	
	1965 Initial	1967 Final	1965 Initial	1967 Final	1965 Initial	1967 Final	1965 Initial	1967 Final
Transfer								
M	14.86	15.45	7.95	9.64	9.67	12.77	23.35	15.01
SD	3.31	2.84	2.14	2.69	3.24	4.56	6.69	9.52
n	52	59	54	61	52	59	55	62
Engineer								
M	14.27	14.57	7.01	7.17	8.18	9.19	26.43	22.2
SD	2.81	2.84	2.52	2.64	2.84	3.89	6.28	9.12
n	141	167	143	168	139	167	142	164
Estimate of ω^2 for difference between means	0	1.47%*	2.42%*	14.45%**	4.31%*	13.12%**	3.98%*	10.81%**

* *t* for difference between engineer and transfer means significant at $p < .025$.
** $p < .001$.

RESULTS

The results of the statistical operations do not supply consistent, unequivocal support for the hypotheses in this study. Table 1 presents the index means and standard deviations for the transfer and engineer portions of the holdout sample. In virtually every case the trend of the mean scores over time indicates an increase (decrease in the case of authoritarianism) in the same direction for both transfers and engineers.

Hypothesis 1 stated that there would be rather small (or no) differences between engineers and transfers with regard to their prefreshman (initial) scores on the four indexes. This is generally true with the average difference among the initial wave means equal to only about 1.5 points although 3 out of 4 of the differences are statistically significant. Moreover, the expectation that these differences would be small even though significant is supported by the estimate of ω^2, which does not exceed 5% for any of the initial differences. The statistic, ω^2, is a proportional reduction in error measure indicating the strength of association (see Costner, 1965; Hays, 1963).

When sophomore (final) scores on the four indexes in Table 1 are compared (i.e., transfers versus engineers), the results are strikingly different. In every case there is a substantial increase in ω^2, and in the confidence level at which the null hypothesis may be rejected.

The degree of accentuation shown on the indexes is not very great—with the exception of the authoritarianism index. Table 2 shows the mean change scores for the criterion groups. Hypothesis 3, which stated that change scores would be greater for transfers, and the estimated ω^2 would be greater for change scores than for either initial or final scores, is partly supported by these data. With the exception of the friendship index, the differences in means provide support for the hypothesis which suggests that change scores are associated with attrition. Support is not found, however, for the hypothesis that an index change score accounts for a greater amount of criterion variable variance than either the initial or final score on that index.

TABLE 2

CHANGE SCORE MEANS BY CRITERION GROUPS

Group	Index			
	Friendship	Self-perception	Educational interest	Authoritarianism
Transfers				
M	.35	1.81*	2.97*	−8.66*
SD	4.11	2.28	4.41	9.78
n	50	53	50	55
Engineers				
M	.22	.06	.92*	−4.58*
SD	3.67	2.64	4.06	7.08
n	139	142	138	137
Estimate of ω^2 for difference between change score Ms	0	9.15%**	4.08%**	4.74%**

* change score means different from zero at $p < .005$.
** *t* for difference between engineer and transfer change score means significant at $p < .005$.

In order to ascertain the number of sub-

TABLE 3

Change Frequencies by Criterion Groups

Group	Index			
	Friendship	Self-perception	Education	Authoritarianism
Number of engineers				
Increasing	72 (53%)	65 (46%)	76 (55%)	29 (21%)
Decreasing	67 (48%)	77 (54%)	62 (45%)	108 (79%)*
Number of transfers				
Increasing	28 (56%)	41 (77%)	38 (76%)	6 (11%)
Decreasing	22 (44%)	12 (23%)*	12 (24%)*	48 (89%)*

* Intragroup differences significant at $p < .001$.

jects whose scores changed, each subject's sophomore index score was examined to determine whether it had increased, decreased, or remained the same relative to his prefreshman score. An increase in the sophomore score relative to his prefreshman score does not necessarily indicate accentuation, since accentuation applies only to an increase (or decrease) in scores which were initially high (or low). Table 3 does, however, indicate the number of subjects moving in each direction relative to their own initial score.

If chance alone was operating, one would expect 50% of the subjects to increase and 50% to decrease on each index. Within the transfer group significant deviations from the chance distribution occur on three out of four of the indexes. These results are consistent with Hypothesis 2 and 3.

As predicted by Hypothesis 4 the socialization index discriminated between transfers and engineers at an adequate level ($p < .001$) but it yielded only a 7.16% reduction in error.

Hypothesis 2 stated that engineers' scores would be more stable over time than transfers'. The coefficient of correlation, r_{if}, in Table 4 refers to the relation of prefreshman (i = initial) index scores to sophomore (f = final) scores. With the exception of the test-retest coefficient for the friendship index based on the transfer sample ($r_{tif} = .062$), all the correlations in Table 4 are significantly different from zero at $p < .05$. Only in the case of the authoritarianism index, however, is the stability coefficient for the engineers different from (at $p < .007$) that for the transfers. The moderate to low stability coefficients are rather disappointing since they tend to put a statistical limit on validity. The authoritarianism index is the only one which both provides support for Hypothesis 2 and indicates an adequate degree of reliability. It should be noted here that the authoritarianism index was constructed in a fashion slightly different from the other indexes. The items were selected on the basis of their correlation with another test taken in the sophomore administration rather than with the dependent variable of attrition. Since the dependent variable is, itself, both reliable and valid, this would tend to indicate that the reliability of the authoritarianism index is probably a lower-bound estimate. That is, the items for the authoritarianism index were chosen on the basis of a less valid and less reliable criterion than the items for the other indexes, yet it shows the greatest reliability. This is not an especially unlikely phenomenon. Dawes and Meehl (1966) and Cronbach and Meehl (1955) have referred to it as "the bootstraps phenomenon"—a situation in which a test may be more valid

TABLE 4

Index Test-Retest Correlations by Criterion Groups

Group	Index			
	Friendship	Self-perception	Education	Authoritarianism
Transfers				
r_{if}	.062	.543	.421	.320
n	50	53	50	55
Engineer				
r_{if}	.172	.453	.294	.622
n	139	142	138	137

(and/or reliable) than the criterion on which it was validated.

When indexes are constructed using item analysis techniques and an external criterion it is essential that they demonstrate adequate internal homogeneity and desirable that they be independent of each other. Coefficient alpha is a useful and familiar homogeneity measure; for the friendship index $\alpha = .164$, for the self-perception index $\alpha = .493$, for the education index $\alpha = .798$, and for the authoritarianism index $\alpha = .906$. The measures are adequate for all but the friendship index. The reliability of homogeneity measures has been discussed by Athanasiou (1969).

For practical purposes the intercorrelation among the indexes indicate that they are nearly orthogonal. The largest correlation ($r = .25$) is between the self-perception index and the education index, but it indicates only 6¼% shared variance.

Hypothesis 4 stated that the socialization index should predict change scores on the other indexes. Unfortunately, neither the engineer sample, the transfer sample, nor the combined sample yielded data which could support this part of Hypothesis 4. It may be that the linear prediction rule does not do a very good job in summarizing the tendency of change scores to differ systematically with differences on the socialization index (see Hays, 1963, pp. 509–510).

If accentuation has taken place, change scores should correlate with the criterion variable at a higher level when initial position is held constant than when initial position is allowed to vary. Table 5 contains the values of the criterion (C) and change scores (D) correlations, X_1 refers to the initial position on an index which is held constant. The correlations are point-biserial coefficients with transfer status coded "1" and engineer status coded "0."

TABLE 5

PARTIAL CORRELATIONS OF CHANGE SCORES WITH CRITERION

Index	r_{CD}	$r_{CD \cdot X_1}$
Friendship	.022	.10
Self-perception	.293	.41
Education	.218	.315
Authoritarianism	.207	.25

Note.—Abbreviated: C = criterion, D = change score, X_1 = initial position on index. For three variables, $r > .244$ is significant at $p < .01$.

TABLE 6

MULTIPLE CORRELATION OF INDEX SCORES WITH CRITERION BY TESTING WAVES

Score	R	R^2
1965 prefreshman index scores	.29	8%
1967 sophomore index scores	.48	23%
Change scores (with initial position held constant)	.55	30%

Note.—for five variables, $R \geq .29$ is significant at $p < .01$.

To assess the value of the indexes as multiple predictors of the criterion, several composite correlations were computed. Table 6 contains the multiple Rs and R^2, the fraction of variance explained.

The interesting thing to note in Table 6 is the increasing fraction of variance accounted for, over time, by the multiple correlation coefficient. For freshman scores, R^2 is low, it increases by nearly a factor of 3 to 23% by the sophomore testing. Change scores (with initial index positions held constant) produce a multiple R which accounts for 30% of the variance. These findings are in line with Hypotheses 1 through 3.

DISCUSSION

Hypothesis 1 stated that there would be little difference between transfers and engineers on entrance characteristics. Relative to their exit (sophomore) differences this is true. On the other hand, the multiple correlation of the prefreshman index scores with the criterion is .29, and is significant at $p < .01$. The difference between the engineers and transfers at entrance is certainly "real," but it is hard to say whether or not it is a nontrivial difference. Given the complexity of individual personality and the college environment, one might consider that accounting for 8% of the criterion variable variance with responses collected nearly 18 months in advance of criterion measurement is nontrivial accomplishment.

Hypothesis 2 stated that engineers would show greater stability of scores than transfers. This was true only in the case of the authoritarianism index. For the other indexes the test-retest reliability coefficients were essentially the same for engineers and transfers. Hypothesis 2 also stated that there would be greater differences between engineers and transfers (on the four indexes) at exit than there was at entrance. The data in Table 2 clearly support this hypothesis. The proportion of criterion variance accounted for by each index (except friendship) almost triples and in one case quadruples over the period between pre-freshman and sophomore administrations. The data also indicate that most of this increase in validity is due to increases in index scores on the part of transfers.

Hypothesis 3, which stated that change scores on the indexes would be greater for transfers and that change scores would be more valid than either entrance or exit scores, received support since both frequency and amount of change was predictive of transfer status. The multiple correlation of change scores with the criterion variable was greater (with initial position held constant) than the multiple correlation of either the initial or final scores with the criterion. The difference, however, was rather small—only seven percentage points—between the amounts of criterion variance accounted for by change scores and by final scores.

For each index the correlation of change score with criterion increases when initial position is partialed out. The heuristic value of these findings lies in their indication that change is not wholly a function of initial position.

The comparison of entrance versus exit scores, the size of change scores, the frequencies of change, the partial correlations, and the coefficients of multiple determination clearly show that over time, on the indexes used in this study, transfer students increase the initial gap between themselves and engineers. These findings may be added to those adduced by Feldman and Newcomb (1969) and provide the suggestion that accentuation can take place within a major field (engineering) even when only small initial differences are present.

The greatest problem in this study is the lack of support for Hypothesis 4 which suggested that the socialization index would provide an indication of the degree of change experienced by the subjects. This socialization index functioned adequately as a dependent variable, but not at all as an intervening variable, which would predict valid change scores.

Virtually every legitimate means of testing the validity of the socialization index and its items was attempted. For example, each item was tested against change scores. This produced nearly 100 point-biserial correlations, of which only 8 were significant at $p < .05$. There was no discernible pattern to these correlations and they were all rather low.

It may be that the wrong items were chosen, but they do appear to have face validity. It may be that they were scored wrong, but the same cross-validation procedures were used with these items as with items for the other indexes. It may be that the items were the "right" ones, but they predict to different kinds of change not measured here. Perhaps the decision to change is made more on a psychological basis than a social-psychological one. That is, the decision-making process may involve processes more internal than those measured by the questionnaire. If the decision were made through extensive introspection and withdrawal from the social structure, the questionnaire would be relatively insensitive to such processes. Direct in-depth interviews might ferret out such changes, but there was neither time nor money available to pursue this approach.

There is now clear evidence for accentuation both within and between academic major fields. The task which remains is to ferret out the etiology of these changes. From these data it appears that the socialization process (via membership groups, etc.) is not correlated with change.

The problem of high attrition rates from the study of engineering also remains and these data seem to indicate that the problem will not be solved by either early selection programs (cf. Athanasiou, 1968b, for a discussion of entrance differences on Scholastic Aptitude Tests, etc.) or the manipulation of group socializing techniques (dormitory groupings, clubs, etc.).

Perhaps an answer to reducing the high attrition rate among engineering students lies in permitting and encouraging the expression of diverse attitudes through the curriculum. That is, greater tolerance of diversity within a field might lead to decreasing diversity between fields. A change of this sort would at least require a change in the structure of the curriculum. It is likely that some change in the content of the curriculum would also be required. For example, rather than chart out a student's entire classroom career from his first to last semester, the curriculum might aim at providing a broad general education in the first 2 years. This would allow the student to sample courses from the engineering school departments and give him some time to decide on his specialty. The subsequent 3 years could be devoted to focused work toward a master's degree.

It seems unlikely that the various departments within an engineering school would give up their hold (via curricular requirements) on the students without some struggle. In the interim, counselors for both freshman and sophomores should be made aware of the attitudinal differences and curriculum preferences among students.

REFERENCES

Athanasiou, R. B. Reliability and validity of a predictor of engineering student attrition. *Proceedings of the 76th Annual Convention of the American Psychological Association,* 1968, **3,** 171–172. (a)

Athanasiou, R. B. Technique without mystique: A study of authoritarianism in engineering students. *Educational and Psychological Measurement,* 1968, **28,** 1181–1188. (b)

Athanasiou, R. B. The reliability of homogeneity. *Proceedings of the 77th Annual Convention of the American Psychological Association,* 1969, **4,** 119–120.

Beal, L., & Bordin, E. S. The development and personality of engineers. *Personnel and Guidance Journal,* 1964, **43,** 23–32.

Bereiter, C. Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change.* Madison: University of Wisconsin Press, 1967.

Brehm, J., & Cohen, A. *Explorations in cognitive dissonance.* New York: Wiley, 1964.

Brim, O. G., Jr., & Wheeler, S. *Socialization after childhood.* New York: Wiley, 1966.

Costner, H. L. Criteria for measures of association. *American Sociological Review,* 1965, **30,** 341–353.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. *Psychological Bulletin,* 1955, **52,** 281–302.

Dawes, R., & Meehl, P. Mixed group validation: A method for determining the validity of diagnostic signs without using criterion groups. *Psychological Bulletin,* 1966, **66,** 63–67.

Feldman, K. A., & Newcomb, T. M. *The impacts of colleges upon their students.* New York: Jossey-Bass, 1969.

Guilford, J. P. *Psychometric methods* (2nd ed.). New York: McGraw-Hill, 1954.

Hays, W. L. *Statistics for psychologists.* New York: Holt, Rinehart, & Winston, 1963.

Kanter, S. A. The social psychology of premature occupational choice. Unpublished doctoral dissertation, University of Michigan, 1967.

Keniston, K. *The uncommitted.* New York: Harcourt Brace & World, 1965.

Kulberg, F. E., & Owens, W. A. Some life history antecedents of engineering interests. *Journal of Educational Psychology,* 1960, **51,** 26–31.

Mosier, M. F., & Kuder, G. F. Personal preference differences among occupational groups. *Journal of Applied Psychology,* 1949, **33,** 231–239.

Newcomb, T. M. *Personality and social change.* New York: Dryden, 1943.

Sanford, N. (Ed.) *The American college.* New York: Wiley, 1962.

Sanford, N. *Self and society.* New York: Atherton, 1966.

Sanford, N. *Where colleges fail.* Jossey-Bass, 1967.

Swap, W. C. The accentuation of values. Unpublished paper, the University of Michigan, 1968.

White, R. W. *Lives in progress.* New York. Holt, 1952.

(Received February 19, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 2, 167–178

TRANSFER FROM TEACHING TO LEARNING[1]†

K. KILBY LONG[2]
University of New Mexico

Serving as an experimenter in a paired-associate learning procedure was observed to have a beneficial effect on subsequent learning. Twenty subjects in an initial experiment, and 105 subjects in an expanded replication were used to investigate this effect which is presented as a transfer from teaching to learning. A number of control conditions failed to produce any postexperimental effect; a finding which suggested that the specific "teaching" experience did account for the effect which was observed. In particular, one control which required teaching silent subjects failed to produce a reliable effect on subsequent learning. This finding is discussed in terms of the importance of learning to utilize feedback in the teaching-learning process.

In a recent review of instructional psychology, Gagné and Rohwer (1969) cite the commonly held belief that a good way to learn material is to teach it. The general notion that the experience of teaching has beneficial effects on the learning of the teacher is widely accepted, and may also be seen in some suggestions for student teaching. McPhie (1969), for instance, suggests that student reports in social studies classes are of potential benefit to the student reporting as well as to the class; and Amidon and Hunter (1967, p. 82) state that allowing children to participate in the dissemination of information is a good technique for instituting verbal interaction, and may result in improved classroom motivation. Unfortunately, such suggestions for student presentations in the classroom have been given without any clear idea of what the specific benefits of such activity may be. Gagné and Rohwer, in fact, question the whole idea of learning material by teaching. Research is currently needed to demonstrate that there is an empirical basis for having students teach as an instructional procedure.

It may be that the beneficial effects of teaching, in terms of the learning of the teacher, are not in the acquisition of the material being taught, but rather in the learning of subsequent material. In other words, there may be a transfer from an experience an individual has in teaching some skill or task to the subsequent learning of a related skill or task. The idea of a specific transfer from teaching to subsequent learning has been recognized by Ellis (1968, pp. 204–205) as an application of transfer principles to the educational context. Ellis contends that the individual serving as a teacher may be learning something about the complexities of the stimulus environment; he may be learning how to attend.

The present research is addressed to the specific question of a possible transfer from teaching activity to subsequent learning. Two experiments were conducted in order to determine if the effect of teaching on subsequent learning could be demonstrated within the confines of a formal laboratory learning procedure.

There is some indirect evidence for a transfer from teaching to learning based on peer-tutoring reports. Cloward (1967) reports a study of tutoring experiences given students in New York City which obtained data showing a beneficial effect of tutoring on the reading ability of the tutors.

Lippitt and Lippitt (1968) report that older students, tutors, gain valuable social and academic benefits from helping younger children learn. Similarly, D. Willis (personal communication, 1968) reports such an effect in high school students in Albuquerque, New Mexico, who were given experience tutoring other younger children. The Albuquerque tutors were reported to have shown improvements in academic performance as well as in general classroom conduct. In fact, Willis cites the peer-tutoring experience as a method of bolstering the classroom motivation of possible school dropouts.

---

[1] This report is based on a thesis submitted in partial fulfillment of the requirements for the doctoral degree at the University of New Mexico, 1970, under the direction of Henry C. Ellis, to whom the author would like to express his gratitude. Appreciation is also directed to Frank A. Logan, Douglas P. Ferraro, David T. Benedetti, and David W. Bessemer, members of the authors' dissertation committee, for their assistance.

[2] Due to the death of K. Kilby Long, requests for reprints should be sent to Ronald W. Shaffer, Department of Psychology, Western Washington State College, Bellingham, Washington, 98225.

Unfortunately, the reports of peer tutoring are not precisely transfer studies; nor have the observations of the effects of tutoring on the tutors been made with the controls necessary for the analysis of the source of improvement in the tutors' behavior. Such reports are only suggestive until better data are obtained.

Another general type of literature which suggests the idea of a transfer from teaching to learning is that of role playing. A number of authors have used role playing in the classroom as a procedure for imparting beneficial effects to those playing a variety of roles (Kuethe, 1968, p. 133; Mann, 1956; Pressey, Robinson, & Horrocks, 1959, pp. 562–563; Zander & Lippett, 1944). Pressey et al., for instance, feel that having the student play the role of the teacher might provide the student an opportunity to learn from the reactions of other students in the class to his own "teacher" activity. Kuethe, as well, has specifically recommended playing the role of the teacher in the classroom as an instructional activity.

It should be pointed out that the statements from the role-playing literature are of a speculative nature, and, furthermore, are not based on firm data. Nevertheless, both the role-playing literature and the peer-tutoring literature are suggestive of an effect of teaching on subsequent learning; and both types of literature are certainly encouraging for consideration of further research into the effect.

A study in the area of incidental learning, reported by Myers, Travers, and Sanford (1965) used an experimental procedure which is appropriate to the study of a possible transfer from teaching to learning. Myers et al. ran what amounts to a paired-associate learning procedure in which student pairs were used. Each student pair consisted of a student experimenter who presented stimulus materials and feedback, and a student subject who attempted to learn the material. The authors were interested in comparing the degree of incidental learning, defined as the amount of stimulus material recalled by the student experimenters on a subsequent recall task, with the recall data of the student subjects which was defined as a measure of direct learning. As in the case of earlier experiments in incidental learning, which measured the amount of material learned by the experimenters as they presented that material (e.g., Jenkins, 1933; Postman & Phillips, 1954), Myers et al. (1965) found incidental learning to be inferior to direct learning. The student experimenter group made considerably fewer correct responses on the criterion task than did the student subject group. Unfortunately for the notion of a transfer from teaching to learning, Myers et al. did not require any subsequent performance of their student experimenters.

The two experiments of the present research were primarily intended to provide an empirical demonstration of the transfer from teaching to learning. The research utilized a student-pair procedure which was quite similar to the Myers et al. procedure. "Teaching," in this case, is defined as the activity required to present stimulus materials, before and following subjects' responses during the learning task.

Both experiments utilized a three-stage design which was built around a series of 12 paired-associate learning lists. The general learning procedure and stimulus materials were adapted from experiments by Rogers (1967) and Long (1969). In the first stage, the initial lists acquisition stage, four of the 12 paired-associate lists were presented for two anticipation trials each to all subjects. In the second stage, the experimental stage, various tasks were imposed on the different groups of the two experiments. The third stage, the transfer lists, required that all subjects learn the final four paired-associate lists in the same fashion as Stage 1 of the design. Differences between the various conditions in the changes in performance from Stage 1 to Stage 3 were regarded as indicative of experimental effects. Data from both of the anticipation trials of each list plus data from a recall task at the end of each stage were analyzed.

In the first experiment, there were only two groups of subjects. One group served as experimenters, and engaged in teaching activity during Stage 2 of the procedure; the other group served as subjects during all three stages. The purpose of this experiment was to see if the teaching experience given the first group would facilitate the subsequent transfer-task performance relative to that of the second group, which had continued experience learning the material. The second experiment replicated the two groups of the first experiment, and added five additional control conditions. The logic of these new controls will be described subsequently. The purpose of the second experiment was to examine several possible sources of a transfer from teaching to learning within the general confines of the experimental procedures of the first experiment.

EXPERIMENT I

Method

Experimental design. Table 1 presents a description of the conditions of Experiment I. The three-stage, two-group design employed was a mixed design, with trials treated as a repeated measure within subjects, and two conditions as a between measure. Ten subjects were randomly assigned to each two groups. The subjects were volunteers from the Introductory Psychology laboratory classes at the University of New Mexico. All subjects of both groups were randomly assigned a different order of a set of 12 paired-associate lists. The lists consisted of 10 pairs of common English words, and were presented for two paired-associate anticipation trials per list. Each individual sequence of 12 lists was presented in three stages to all subjects. List 1–4 in the sequence were presented during Stage 1, the initial lists acquisition stage. Lists 5–8 were presented during Stage 2, the experimental treatments stage; and Lists 9–12 were presented during Stage 3, the transfer task stage. There was a recall task followed by a 5-minute rest break after each stage.

Stage 1. In the initial lists acquisition stage, all subjects in both groups were presented with four of the paired-associate lists from their individual sequence of 12 lists. The stimulus materials were prepared on 3 × 5 cards. On one side of a card the stimulus member of a pair words was typed in uppercase letters. The stimulus and response members of the pair were both typed on the other side of the card in uppercase letters. The pair of words was also typed upside down in lowercase letters near the upper left-hand corner of the stimulus-response side of the cards so that the experimenter could read the pair of words as he was presenting the materials.

TABLE 1

CONDITIONS OF EXPERIMENT I

Group	Stage 1 (Initial lists acquisition)	Stage 2 (Experimental treatments)	Stage 3 (Transfer lists)
Teach Other	Learn Lists 1–4	Teach Lists 5–8	Learn Lists 9–12
Learn	Learn Lists 1–4	Learn Lists 5–8	Learn Lists 9–12

The procedure for running an individual subject through Stage 1 was as follows:

The subject was instructed that he was participating in an experiment which was designed to investigate a particular type of material, and a particular method for presenting that material. The subject was told that there would be a series of lists of word pairs typed on 3 × 5 cards that he was to try and learn to the best of his ability; and that all lists would be presented in the same fashion. If there were no questions, the first paired-associate list was presented.

The subject was shown the first set of 10 stimulus cards, and told that each card would be presented to him three times. During the first presentation, the subject was simply to read aloud the single word on the front of the card. When the card was turned over, he was to read aloud the pair of words which appeared on the back side. Each card in the first list was presented in this fashion. When the experimenter had finished with the first and second presentation of the list, he shuffled the cards before presenting them again.

For the second and third presentation of the list, the subject was instructed to say the first word aloud immediately upon its presentation, and then to try and say the word that went with it before the experimenter turned the card over and said the pair of words himself. The subject was instructed to repeat the pair of words after the experimenter said them, whether or not the subject had correctly anticipated the response member of the pair.

The experimenter presented each subject with the four lists of Stage 1 in the same manner, moving from one list to the next immediately upon finishing with the preceding list.

Exposure time of both the stimulus and stimulus-response members of the word pairs was 2–3 seconds. This timing was somewhat inexact since all presentation of the stimulus materials was done by hand using the 3 × 5 cards, and some variation in timing was allowed for each subject, according to that subject's ability to handle the procedure of the learning task. After the presentation of the stimulus-response member of a word pair, that card was immediately put down, and the next card picked up and presented to the subject.

Following the completion of the presentation of Lists 1–4 during Stage 1, a recall task was administered. Following the recall task, the subjects were given a 5-minute rest break before moving on to Stage 2.

Stage 2. During the experimental stage, the subjects in the Teach Other group were instructed that the experimenter was interested in their reactions to the stimulus material and the method of presentation from the viewpoint of an experimenter. Teach Other group subjects were told that they would present four lists of word pairs to a new subject. Each of the 10 Teach Other group subjects then presented the stimulus materials of Lists 5–8 plus feedback to one of 10 new subjects in exactly the same fashion as the presentation of Lists 1–4.

The Learn group was presented Lists 5–8 just as they were presented Lists 1–4.

Following the presentation of Lists 5–8, both groups of subjects plus the new subjects for the Teach Other group were given a recall task. A 5-minute rest break again preceded the beginning of the next stage.

Stage 3. The transfer task stage consisted of presenting both original groups of subjects with Lists 9–12 in exactly the same fashion as in Stage 1. A recall task following Stage 3 concluded the experimental procedure.

Recall Task. After each stage of list presentation, whether taught or learned, all subjects were immediately given a recall task. This task consisted of giving the subject four sheets of paper on which the stimulus members of each of the four preceding paired-associate lists were printed, and asking them to write in as many of the associates to the printed words as they could remember. Five minutes were allowed to work on the recall task.

RESULTS

Table 2 represents the mean number of correct anticipation responses for both anticipation trials obtained from the initial lists acquisition and the transfer task stages of the experiment. The mean changes from Stage 1 to Stage 3 are also presented in Table 2. These change scores represent the simple numerical difference between the performance on Lists 9–12 and the initial performance on Lists 1–4. A positive score indicates improvement following the experimental treatments during Lists 5–8; and a negative score represents a performance decrement following the experimental treatments.

Stage 1. Examination of the Stage 1 data in Table 2 points out essentially no differences between the two groups on Trial 1 and Trial 2. Performance during the first stage across the two conditions, however, shows an expected improvement on the second anticipation trial. Analysis of variance of the Stage 1 data revealed no significant differences between the conditions of the experiment, as well as an insignificant interaction between conditions and trials. Accordingly, it may be concluded that performance between the Teach Other and the Learn groups was equivalent on Lists 1–4. A significant F for trials, $F = 19.32$, $df = 1/18$, $p < .01$, was obtained. This simply reflects the improvement on the repeated measure shown by both groups.

Stage 3. Performance on the transfer lists, as indicated by the Stage 3 data, clearly shows a difference, on both trials, between the Teach Other and the Learn groups. This difference was significant, $F = 5.02$, $df = 1/18$, $p < .05$, as was the improvement in performance for both groups from Trial 1 to Trial 2, $F = 23.92$, $df = 1/18$, $p < .01$. The interaction between conditions and trials was not found to be significant.

Differences between the two conditions on the transfer lists, coupled with equivalent performance during initial lists learning, indicate the presence of a transfer effect. This

TABLE 2

ANTICIPATION PERFORMANCE MEASURES IN EXPERIMENT I

Group	Mean correct anticipations		Changes (Stage 1–Stage 3)
	Stage 1 (Initial lists acquisition)	Stage 3 (Transfer lists)	
Teach Other			
Trial 1	14.2	16.9	2.70
Trial 2	20.1	27.6	7.50
Learn			
Trial 1	13.2	13.4	.20
Trial 2	20.9	20.3	−.60

effect is further illustrated with the change score data in Table 2.

Changes. The analysis of simple change scores was undertaken to give a direct picture of the effects of the interpolated activity in terms of the initial performance of the two groups. Since the groups were quite comparable in performance on Lists 1–4, scores representing a change from that initial performance are an appropriate index of an experimental effect for this type of design (Campbell & Stanley, 1968, p. 23).

Analysis of variance performed on the mean changes of the two groups revealed a significant difference between conditions, $F = 9.13$, $df = 1/18$, $p < .01$, as well as a significant interaction between conditions and trials, $F = 10.00$, $df = 1/18$, $p < .01$. This interaction is apparent in the changes of Table 2, and reflects the greater improvement in the Teach Other group on the second anticipation trial of the transfer lists over that seen on the first anticipation trials. The interaction between conditions and trials is an important finding in that it may point to increased effort on the learning task following the teaching activity. The Teach Other group subjects may be utilizing the second anticipation trials of Lists 9–12 to more successfully master that material. A significant main effect for trials, $F = 5.16$, $df = 1/18$, $p < .05$, was also obtained; however, this was thought to reflect primarily the major differences between the changes for Trial 1 and Trial 2 in the Teach Other group. The changes observed in the Learn group are so small that it seems unlikely that the differences between the two anticipation trials represent any meaningful effect.

Recall task. Table 3 presents the mean number of correct recall responses for the two groups. Stage 1, Stage 3, and the changes from Stage 1 to Stage 3 are presented.

Recall performance over the initial lists was equivalent, as can be seen by the Stage 1 performance in Table 3. The differences between the two groups on the transfer lists were reliable, as evidenced by a t of 2.38, $p < .05$. A test of significance between the change score means produced a t of 3.21, $p < .01$.

TABLE 3

MEAN NUMBER OF CORRECT RECALL RESPONSES IN EXPERIMENT I

Group	Stage 1 (Initial lists acquisition)	Stage 3 (Transfer lists)	Changes (Stage 3–Stage 1)
Teach Other	16.40	22.80	6.20
Learn	16.30	16.60	.30

The experimental effect observed in the correct anticipation performance between the Teach Other group and the Learn group was found to have occurred on the recall task as well. This finding indicates that the differences between the two groups represent the acquisition of more of the Lists 9–12 material by the Teach Other group, as evidenced by subsequent superior recall; rather than a temporary performance effect, limited to the anticipation data. The fact that recall performance improved, just as anticipation performance during learning improved, is evidence for a strong experimental effect. The findings of Experiment I encouraged a second experiment which attempted to attribute the differences between the two groups of the first experiment specifically to the teaching experience.

EXPERIMENT II

The results of Experiment I suggested an improvement in learning following a teaching experience. Unfortunately, the differences between the Teach Other and the Learn group of the first experiment are not sufficient, by themselves, to conclude that the experience presenting Lists 5–8 to a new subject produced the improvement in subsequent performance observed in the Teach Other group. Experiment II replicated the Teach Other and the Learn groups of Experiment I, and added five new conditions. The purpose of the new conditions was to attribute such transfer to the teaching activity engaged in by the Teach Other group during Lists 5–8, and to provide some information about the nature of the observed improvement.

It is conceivable that the differences in Experiment I between the Teach Other and the Learn groups were due not to the teaching activity per se, but rather to the additional learning of Lists 5–8 required of the Learn group during Stage 2 of the experiment. In other words, there may have been a proactive inhibition effect of the first eight lists which would depress performance on the transfer task for the Learn group. The first new condition in Experiment II was a control for a possible detrimental effect on Lists 9–12 of learning Lists 5–8. This control group simply rested for a period of time equivalent to that necessary for subjects to learn Lists 5–8, before learning the lists of the transfer task. Should performance on the lists following the extended rest break show performance similar to that of the Learn group, support would be given to the contention that the differences between the Teach Other and the Learn groups were due to the teaching activity of the Teach Other group, and would minimize any interference-reduction interpretation.

A second control condition was one which simply required subjects to sort cards in a particular manner similar to the stimulus materials of the other groups during the period when Lists 5–8 were presented. This group was a control for possible warm-up effects of the activity required of the Teach Other group subjects during Lists 5–8, and provided a control for any contention that improvement observed in the Teach Other group was due to activity alone.

The remaining controls were all addressed to the belief that subjects are learning something as they teach which then facilitates their own subsequent learning. It may be recalled that this idea was formulated by Ellis (1968) in terms of learning about the stimulus environment while teaching, and is reflected by the suggestions for learning while role playing (Kuethe, 1968, p. 133; Pressey et al., pp. 562– 563). In the present case, several possibilities for learning while teaching were explored by a group which taught themselves, a group which simply observed the teaching–learning process, and a group which taught subject who did not respond aloud.

It is conceivable that the improvement observed in the Teach Other group is an effect of some sort of self-teaching of the materials of Lists 5–8, regardless of any effect of the personal interaction with a new subject which occurred as the Teach Other group presented stimulus materials and feedback. Through teaching, it may be possible to learn about the task at hand by being involved with that task in a way other than that of the learner. A control group which presented the materials of Lists 5–8 to themselves was included in order to examine the possibility that the improvement in learning following teaching is due to self-teaching while presenting material to someone else.

It is also possible that the improvement following teaching is due to learning about the entire teaching–learning process while presenting material to an actively responding subject. To examine this possibility, a control was included which simply required subjects to observe the presentation of Lists 5–8 to new subjects. Perhaps a teacher learns to work better within a teaching–learning procedure by being able to watch someone else learn within that procedure. If this is the case, the same possibility is present for someone who observes the procedure without participating in any other way. In fact, it may even be a better opportunity to learn about the teaching–learning procedure, as the observer is not required to attend to the presentation of the material, etc.

Finally, it is conceivable that the improvement observed in the Teach Other group is due to something that group learned from their own subjects' responses to the materials of Lists 5–8. Learning while teaching may involve learning about the behavior of the students. The last additional control was addressed to the possibility of learning from the new subjects' responses to Lists 5–8, and required that this group present the materials of the teaching task to new subjects who never said a word. If an improvement follows teaching nonresponding subjects, it is reasonable to assume that the improvement observed in the Teach Other group is not due to learning from the behavior of that group's "students." If no improvement follows teaching nonresponding subjects, it is suggested that the improvement following the Teach Other group's teaching reflects learning about the teaching–learning process from the Lists 5–8 subjects' behavior.

Method

Experimental design. Experiment II used the same three-stage, mixed design as in the first experiment. The seven groups of Experiment II are schematized in Table 4. Volunteers from the Introductory Psychology lecture classes at the University of New Mexico were randomly assigned to each of the seven groups. There were 15 subjects per group, plus an additional 15 subjects per group for the Teach Other, the Observe, and the Teach Nonresponder groups to use during Stage 2 of the experiment.

Stage 1. All subjects were run through exactly the same presentation of Lists 1–4, plus the recall task, as in Experiment I.

Stage 2. The Lists 5–8 procedure for the Teach Other and the Learn groups was the same as in Experiment I.

The subjects in the Rest group were instructed that they would be taken out of the experiment for a period of time. Those subjects were then taken completely out of the laboratory area and seated before a table with magazines on it. The subjects were allowed to read if they wanted to, or they simply sat at the table and waited. After 20 minutes, approximately the time required to present Lists 5–8 and the recall task, the Rest group subjects were brought back into the experimental room for Stage 3 of the experiment.

The Sort group was instructed that the experimenter was interested in examining a particular type of motor activity related to handling the stimulus cards. These subjects were then shown a new deck of cards which had either the number 1, 2, or 3 typed on the back of each card. The subjects were told to shuffle the cards thoroughly and then to sort the deck, card by card, into three piles corresponding to the numbers. Sorting was done by placing the deck face down on the table, picking up each card individually and turning it over, and placing each card in the appropriate pile. The deck of cards was shuffled and sorted four times, which provided the same number of card manipulations as that provided by the Lists 5–8 procedure for the Teach Other group. A 5-minute rest break preceded Stage 3 for the sort group.

The Stage 2 procedure for the Teach-Self group differed from that for the Learn group only in that the Teach-Self subjects presented the stimuli of Lists 5–8 to themselves as the initial experimenter sat by the side and recorded responses. The Teach-Self group was instructed that the experimenter was interested in the comparison of teaching one's self with being taught by another person. A recall task followed by a 5-minute rest break preceded Stage 3 activity.

The Observe group was instructed to simply watch the general learning process of the paired-associate procedure during Stage 2. They were told that the experimenter was interested in their reactions from a different standpoint. After the rest period following Lists 1–4, the new subject and the original Observe group subject were brought back into the experimental room. Lists 5–8 were administered to the new subject. Both subjects were then given the recall task.

TABLE 4

SCHEMATIZATION OF CONDITIONS OF EXPERIMENT II

Group	Stage 1 (Initial lists acquisition)	Stage 2 (Experimental treatments)	Stage 3 (Transfer lists)
Teach Other	Learn Lists 1–4	Teach Lists 5–8	Learn Lists 9–12
Learn	Learn Lists 1–4	Learn Lists 5–8	Learn Lists 9–12
Rest	Learn Lists 1–4	No activity	Learn Lists 9–12
Sort	Learn Lists 1–4	Sort cards	Learn Lists 9–12
Teach Self	Learn Lists 1–4	Teach Self Lists 5–8	Learn Lists 9–12
Observe	Learn Lists 1–4	Observe Lists 5–8 procedure	Learn Lists 9–12
Teach Nonresponder	Learn Lists 1–4	Teach nonresponding subject Lists 5–8	Learn Lists 9–12

The final group, that which presented Lists 5–8 to silent subjects during Stage 2, did exactly the same thing as the Teach Other group. In this case, however, the new subject never said a word. Instructions to the Teach Nonresponder subjects were the same as to the Teach Other group; that is, that the experimenter was interested in the subject's reactions to the mode of presentation. The new subjects were told simply to study the material for a later recall test. The Teach Nonresponder subjects then presented each of Lists 5–8 one time silently and two times reading the pair of words on each card aloud as they turned the cards over. Following the presentation of the lists, both subjects were given the corresponding recall task.

Stage 3. Lists 9–12 were presented to all the subjects exactly as in Experiment I.

Recall task. These procedures were exactly the same as in Experiment I.

Results

Table 5 presents the mean number of correct anticipation responses for both anticipation trials obtained from Stage 1 and Stage 3, plus the changes from Stage 1 to Stage 3.

Stage 1. Examination of the Stage 1 data in Table 5 points out essentially no differences between any of the groups on Trial 1 or Trial 2. This observation was confirmed by an analysis of variance of the Stage 1 data which revealed no significant differences between any of the conditions, $F < 1.00$, $df = 6/98$, and which produced an insignificant interaction between conditions

TABLE 5

ANTICIPATION PERFORMANCE MEASURES IN EXPERIMENT II

Group	Mean correct anticipations		Changes (Stage 1–Stage 3)
	Stage 1 (Initial lists acquisition)	Stage 3 (Transfer lists)	
Teach Other			
Trial 1	14.33	18.80	4.47
Trial 2	20.67	28.00	7.33
Learn			
Trial 1	13.93	15.66	1.73
Trial 2	23.07	22.40	−.67
Rest			
Trial 1	14.27	13.27	−1.00
Trial 2	22.33	21.87	−.46
Sort			
Trial 1	13.20	13.40	.20
Trial 2	22.07	22.67	.60
Teach Self			
Trial 1	14.33	14.87	.54
Trial 2	21.67	24.53	2.86
Observe			
Trial 1	13.33	13.20	−.13
Trial 2	21.40	22.66	1.26
Teach Nonresponder			
Trial 1	13.53	14.26	.73
Trial 2	20.53	22.67	2.14

and trials, $F < 1.00$, $df = 6/98$. The expected improvement across trials was again found to be reliable, $F = 88.53$, $df = 1/98$, $p < .01$. It may be concluded that, as in Experiment I, performance on the preexperimental treatment lists was the same for all groups.

Stage 3. Examination of the Stage 3 data in Table 5 points out a marked difference between the Teach Other group and all the others on both anticipation trials. Table 5 reveals that the Teach Other group's performance on the transfer lists is superior to that of the other groups in every case. Analysis of variance performed on the Stage 3 data revealed a significant difference between the conditions of the experiment, $F = 4.65$, $df = 6/99$, $p < .01$, and confirmed the improvement in performance across the anticipation trials, $F = 72.56$, $df = 1/98$, $p < .01$.

Changes. As in the case of Experiment I, the clearest indication of the effects of the experimental treatments of Stage 2 is seen in the changes in performance from Stage 1 to Stage 3. In Table 5, the mean changes for the Teach Other group are quite obviously far greater than the changes for any other group, both on Trial 1 and on Trial 2. The differences between the various conditions were reliable, $F = 4.22$, $df = 6/99$, $p < .01$, and the interaction between conditions and trials was also reliable, $F = 3.45$, $df = 6/99$, $p < .01$. This significant interaction illustrates, as in Experiment I, the increased improvement in performance of the Teach Other group on the second anticipation trial over that of the first anticipation trial.

Following the initial analysis of variance tests, Duncan's multiple-range tests (Edwards, 1960, pp. 136–144) were performed on the differences between all possible pairs of mean change scores on Trial 1 and Trial 2. The changes of the Teach Other group were found to differ significantly from those of all six of the other groups on both Trial 1 and Trial 2, and no significant differences were obtained between any of the remaining groups.

The changes of the Teach Other group differed reliably from those of the Learn group ($p < .05$) on both trials, a finding which essentially replicates the general findings of Experiment I. Furthermore, the magnitude of the differences between the mean changes of the two groups (2.74 on Trial 1, 8.00 on Trial 2) is quite similar to the magnitude of the differences between the mean changes of the two groups in the first experiment (2.50 on Trial 1, 8.10 on Trial 2). Accordingly, it is reasonable to regard the experimental effect reflected by the differences between the Teach Other groups and the Learn groups as the same in both experiments.

A slight decrement in performance following the experimental activity was observed in the Rest group, and this change in performance differed significantly from the changes in the Teach Other group on both anticipation trials ($p < .01$). Since no improvement at all followed the extended rest break given the Rest group during Stage 2 of the experiment, it is reasonable to conclude that the differences between the

Teach Other and the Learn groups do not reflect relatively depressed performance of the Learn group on the transfer task due to a proactive inhibition effect of the preceding eight successive lists experienced by that group.

The Sort group also shows essentially no changes between Stage 1 and Stage 3, and the differences between the Sort and the Teach Other group were significant on both trials ($p < .05$ on Trial 1, $p < .01$ on Trial 2). The lack of any effect in the Sort group rules out consideration of a simple "warm up" effect as the source of the observed improvement in the Teach Other groups.

The Teach Self group appears to show some improvement on the second anticipation trial. However, the differences between the changes of the Teach Other and the Teach Self group were significant on both trials ($p < .05$ on Trial 1, $p < .01$ on Trial 2), and no significant differences were found between the Teach Self group and any of the other groups. Given these particular statistical findings, the improvement observed on Trial 2 of the Teach Self groups cannot be regarded as reliably different from the changes of the other groups; and support is not given to an effect of self-teaching as a source of improvement in the Teach Other group's performance.

The differences between the Observe and the Teach Other group's changes were significant on both trials ($p < .05$ on Trial 1, $p < .01$ on Trial 2), and there is little or no suggestion of any effect of observing the teaching–learning process in the changes of the Observe group.

Finally, there is some suggestion of an improvement following the Stage 2 activity in the changes of the Teach Nonresponder group. However, the differences between the changes of the Teach Other group and the Teach Nonresponder group and all of the remaining groups were significant on both trials ($p < .05$ on Trial 1, $p < .01$ on Trial 2). Also, the differences between the Teach Nonresponder group and all of the remaining groups were not significant. The lack of any effect in the Teach Nonresponder group points to some feature of the overt responses of the learners during Lists 5–8 as a possible factor contributing to the improvement in the Teach Other groups.

TABLE 6

MEAN NUMBER CORRECT RECALL RESPONSES IN EXPERIMENT II

Group	Stage 1 (Initial lists acquisition)	Stage 3 (Transfer lists)	Changes (Stage 3–Stage 1
Teach Other	15.33	21.13	5.80
Learn	15.78	15.40	−.38
Rest	14.66	15.40	.74
Sort	15.33	16.20	.87
Teach Self	16.00	18.73	2.73
Observe	13.86	15.27	1.41
Teach Nonresponder	14.80	16.60	1.80

Recall Task. Table 6 shows the recall performance on the lists preceding and following the experimental treatments. The mean changes from Lists 1–4 to Lists 9–12 are also presented in Table 4.

Analysis of variance revealed no significant differences during Lists 1–4; however, significant differences were found between conditions during Lists 9–12, $F = 2.24$, $df = 6/98$, $p < .05$. The analysis performed on the change scores also revealed a significant difference between conditions, $F = 2.61$, $df = 6/98$, $p < .05$.

Tests of significance were run using change scores between all possible pairs of groups. In recall, as in the case of the learning data, there were no significant differences between any pairs except those involving the Teach Other group. The Teach Other group differed significantly from the Observe, Teach Nonresponder, Sort, Rest ($p < .05$), and Learn groups ($p < .01$). The differences between the Teach Other group and the Teach Self group were not found to be significant.

The recall data, in general, are quite similar to the learning data. The Teach Other group's recall performance on Lists 9–12 shows an improvement over that group's performance on Lists 1–4. The Teach Other group's recall performance on Lists 9–12 is also clearly superior to that of the other groups on those lists, with the exception of the Teach Self group in Figure 8. The Teach Self group did not differ statistically from the Teach Other group on the recall measure. Analysis of variance of the individual lists of Stage 3 revealed a significant improvement across Lists 9–12 for all groups, $F = 16.85$, $df = 3/294$, $p < .01$, but failed to find a significant interaction between the lists and the conditions. While the multiple comparisons performed on the recall change found the Teach Other and the Teach Self groups to be similar, the Teach Self group's recall changes did not differ from any of those of the other groups. Accordingly, care should be exercised in making any conclusions about an effect in the Teach Self group.

The recall performance of the Teach Other and the Learn groups of Experiment II parallels that of the two groups in the first experiment. The Teach Other groups in both experiments showed a similar improvement from Stage 1 to Stage 3 (6.20 in Experiment I, 5.80 in Experiment II), and both Learn groups showed essentially no change (.30 in Experiment I, −.38 in Experiment II).

DISCUSSION

The results of the two experiments in the present research demonstrated a beneficial effect of activity defined as teaching on the subsequent learning of those who taught. This finding occurred within the confines of a paired-associate learning procedure in which teaching activity consisted of presenting stimulus materials and response feedback to overtly responding subjects for a period of time. Measures of learning were taken preceding and following the experimental activity of various groups in the two experiments, and a marked improvement was observed to occur for those groups who were given the teaching experience. A number of control conditions failed to show any reliable change in performance on the transfer task measures.

Within the Teach Other groups, the effect of the teaching experience was observed to occur on both anticipation trials of the transfer task lists, and on the recall task which followed that learning activity. Trial 2 anticipation performance showed a greater improvement than that observed on the first anticipation trials, and the amount of improvement in recall performance paralleled that of the second anticipation trials. The greater improvement seen on the second anticipation trials may reflect the Teach Other subjects' more successful utilization of the opportunity to correct mistakes and rehearse associations provided by the second anticipation trial of each list. The fact that the improvement seen in the recall performance was about the same as on the second anticipation trials suggests that the Teach Other subjects were, in fact, learning more of the Lists 9–12 material. Failure to obtain similar improvements in recall as in anticipation performance would have suggested that what improvement was obtained stemmed from a temporary performance factor of some sort.

The improvement in the Teach Other groups was not seen in the Learn groups of the two experiments. With the exception of a small change on the first anticipation trial in the second experiment, both Learn groups evidenced very little change from Lists 1–4 to Lists 9–12 in either anticipation or recall performance. Essentially, then, no improvement seems to have accrued as a function of just the additional experience with the paired-associate material provided by learning Lists 5–8. This conclusion was bolstered by the lack of changes in the Rest group, a group included in the second experiment to examine the possibility that the differences between the Teach Other and the Learn groups reflected a relative depression in the Lists 9–12 performance of the Learn groups, rather than an improvement in the Teach Other groups' performance. This could occur due to a proactive interference effect of learning Lists 5–8 on the transfer task. The Lists 9–12 performance of the Rest group showed no improvement, and it is possible to conclude that the Learn groups' performance does not reflect a detrimental effect of the Lists 5–8 activity, and that the changes in the Teach Other groups reflect actual improvements in performance.

The lack of changes in the Sort group suggest that the Teach Other groups' improvement reflects more than just a warm-up effect of Lists 5–8 teaching activity. This finding supports the contention that factors of the teaching–learning process of the Teach Other groups' Lists 5–8 activity are responsible for the observed improvement in those groups.

If the two anticipation trials' performance plus the recall performance are considered together, there is some suggestion of improvement in the Teach Self, Observe, and Teach Nonresponder groups. All three of these groups appear to be making some improvement on the transfer task. If this is the case, then it may be that the activities required of these three groups which are common to the activity required of the Teach Other groups account, in some part, for the observed improvement in the latter group. However, in no single instance were the changes in the Teach Self, Observe, and Teach Nonresponder groups found to differentiate reliably those groups from any others except the Teach Other group. Accordingly, one may not account for the improvement in the Teach Other groups' performance in terms of the effects of self-teaching, observing the teaching–learning process, or simply presenting material to another person.

The experimental activity of the Teach Nonresponder group differed from that of the Teach Other group only in that the former group presented material to, or taught, subjects who made no overt responses. The Teach Nonresponder group, therefore, did not have a chance to present stimulus material following the responses of the learners, as did the Teach Other group. The Teach Nonresponder group, then, did not have a chance to present "feedback" to their learners. The second and third presentation of each stimulus-response pair of words to overtly responding subjects may be seen as instances of feedback (Deese & Hulse, 1967, pp. 454–456) in that those presentations inform the learner about his responses, and may provide reinforcement. The fact that the Teach Nonresponder group did not improve on the transfer task as did the Teach Other groups strongly suggests the importance of the presentation of stimulus material as feedback as the source of the Teach Other groups' observed improvement. While teaching, the Teach Other group subjects may be learning the importance of utilizing the feedback stimuli. As Ellis (1968) suggests, they may be learning to attend. In this case, they may be learning to attend to aspects of the teaching–learning process which facilitate learning, such as utilization of feedback. This sort of effect is further suggested by the increased improvement seen on the second anticipation trials of Lists 9–12 in the Teach-Other groups, those trials in which utilization of the second stimulus presentation could produce the additional improvement. Further research is needed to fully investigate the effects of this aspect of the present research.

The effects demonstrated in the present research are, of course, highly bound to the experimental situation from which the data were obtained. Presenting material of a paired-associate learning procedure is a highly constricted definition of teaching and great care should be taken in generalizing to any actual classroom considerations. Also, there is no reason to expect that a transfer from teaching to learning is restricted to those teaching–learning processes which use feedback. Given different learning situations, different kinds of teaching, different types of material, there are any number of ways teaching might facilitate subsequent learning. Changes in attitudes about the learning situation might result from having taught, and might then produce greater effort in the following learning situation. Such motivational factors were suggested by Cloward (1967) and D. Willis (personal communication, 1968) as being responsible for changes in the peer-tutors' behavior following the tutoring experiences. The present data may reflect such changes as well, since increased effort on rehearsing the associates after the first anticipation trials may, to some degree, be responsible for the improvement in the Teach Other groups. In terms of learning about features of the teaching–learning process, there are certainly many possibilities beyond learning to use feedback. Strategies of responding, rate of responding, types of errors, and types of correct responses are all examples of what might be learned while teaching.

Further research should examine various teaching–learning processes to determine additional benefits of a teaching experience on the learning of the teacher, and to analyze those benefits into the appropriate sources of transfer.

REFERENCES

AMIDON, C., & HUNTER, E. *Improving teaching, the analysis of classroom verbal interaction.* New York: Holt, Rinehart, & Winston, 1967.

CAMPBELL, D. T., & STANLEY, J. C. *Experimental and quasi-experimental designs for research.* Chicago, Ill.: Rand McNally, 1968.

CLOWARD, R. D. Studies in tutoring. *Journal of Experimental Education,* 1967, **36,** 62–64.

DEESE, J., & HULSE, S. H. *The psychology of learning.* New York: McGraw-Hill, 1967.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart, & Winston, 1960.

ELLIS, H. C. Discussion of Professor E. J. Gibson's paper on "Perceptual learning" and Professor M. Wittrock's paper on "Three conceptual approaches to transfer of training." In R. M. Gagné & W. Gephart (Eds.), *Learning research and school subjects.* Itasca, Ill.: Peacock, 1968.

GAGNÉ, R. M., & ROHWER, W. D. Instructional psychology. *Annual Review of Psychology,* 1969, **20,** 381–418.

JENKINS, J. G. Instructions as a factor in "incidental" learning. *American Journal of Psychology,* 1933, **45,** 471–477.

KUETHE, J. L. *The teaching-learning process.* Glenview, Ill.: Scott, Foresman, 1968.

LIPPITT, R., & LIPPITT, P. Cross-age helpers. *National Education Association Journal,* 1968, **57,** 24–26.

LONG, K. K. An analytical investigation of instructions designed to elicit test anxiety. Unpublished master's thesis, University of New Mexico, 1969.

MANN, J. H. Experimental evaluations of role-playing. *Psychological Bulletin,* 1956, **53,** 227–234.

MYERS, K. E., TRAVERS, R. M. W., & SANFORD, M. E. Learning and reinforcement in student pairs. *Journal of Educational Psychology,* 1965, **56,** 57–62.

MCPHIE, W. E. Student reports and the social studies. In R. E. Gross, W. E. McPhie, & J. R. Fraenkel, (Eds.), *Teaching the social studies; what, why, and how.* Scranton, Penn.: International Textbook Company, 1969.

POSTMAN, L., & PHILLIPS, L. W. Studies in incidental learning: I. The effects of crowding and isolation. *Journal of Experimental Psychology,* 1954, **48,** 48–56.

PRESSEY, S. L., ROBINSON, F. P., & HORROCKS, J. E. *Psychology in education.* New York: Harper, 1959.

ROGERS, J. G. Input and output speed components of learning-to-learn. Unpublished doctoral dissertation, University of New Mexico, 1967.

ZANDER, A. F., & LIPPITT, R. Reality-practice as educational method. *Sociometry,* 1944, **7,** 129–151.

(Received February 16, 1970)

the times are changing—and so are children

A SYMPATHETIC UNDERSTANDING OF THE CHILD:
Six to Sixteen

by DAVID ELKIND
University of Rochester

The author, a Piagetian with many years' experience as teacher and therapist, writes with insight and understanding about the major aspects of child and adolescent development today. He focuses on children in a social context, particularly children in the schools and the schools' influence on their behavior.

The book is divided into two sections, the first dealing with childhood (ages 6–11), the second with adolescence (ages 12–16). For each age group, he gives brief, informal descriptions of the continuum of development; facets of growth and development; social-cultural change; general characteristics; the world of self, home, community; cognitive development; the age group in relation to adults; and adults in relation to the age group. Examples from real life add interest and humor to these discussions.

Ideally suited for courses in Child Psychology and Educational Psychology. 1971, 160 pp., paperbound

now available from

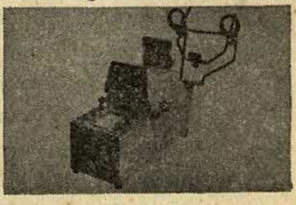
ALLYN & BACON, INC.
College Div., Dept. 893
470 Atlantic Ave., Boston, MA
02210

EASIER, FASTER EYE MOVEMENT RECORDING

eye-trac

The ideal instrument for general purpose eye movement recording featuring simple, economical operation. Eye-Trac permits instantaneous measurement of horizontal eye movements. It is most useful for reading pattern and nystagmus recording. Other applications include measurement of convergence, visual acuity, and binocular fusion. With Eye-Trac you can bypass the time and expense of film processing — recordings are made directly onto heat-sensitive paper which requires no additional processing. Eye-Trac allows even an untrained operator to produce eye movement recordings of high quality.

SPECIFICATIONS:
Weight: 30 lbs.
Maximum Dimensions: 29″ L x 20″ H x 8″ W
Recording Medium: Heat-Sensitive Paper — 2.5″ W — 100′ Roll
Recorder Speed: 10 mm per sec.
Target: Removeable illuminated easel for 3.5 inch cards.
Other targets available.



BIOMETRICS, INC.

INSTRUMENTATION FOR MEDICINE AND BIOLOGY
40 Ames Street Cambridge, Mass. 02142
Tel: (617) 864-9877

JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 62 June 1971 Number 3

Diary No 367
Date 13.9.71
Lib
Bureau Edn'l Psy Research

NOTICE

This journal is participating in a trial of the Current Research Summaries system (CRS) developed by the APA Communications Program. The purpose of the CRS system is to provide a communication channel that will bridge the gap between the time of completing research and formal publication of the findings in the literature. Specifically, it will make the existence of follow-up research known to a larger audience of psychologists, without jeopardizing the publication prerogatives of the author and the "privileged communication" aspects of the informal channels that currently convey this type of information.

CRS solicits information regarding the next step of the research from authors at the time they receive notification that their article has been accepted by this journal. The CRS questionnaire asks for a descriptive title of the work, its purpose, current status, record of prior presentations of findings, and any pertinent references for this latest stage of the research program which did not appear in the original article. Completion of the form is voluntary. If completed, a dagger (†) will appear next to the title of the published article, which indicates that such information regarding the next stage of the research is on file in the Office of Communications and will be sent, in noncitable form, to readers requesting it. All information provided by authors is considered perishable and will be removed from the files after a two-year period.

CRS information is now on file for a number of articles in this issue and the previous issue. These articles are identified by a dagger next to the title. Address your request to: Current Research Summaries, American Psychological Association, 1200 Seventeenth Street, N.W., Washington, D.C. 20036. With each request, please include author's name, article title, journal title, volume and page number. These request instructions will continue to appear on the inside cover of future issues of this journal.

ed .. 179
ng .. 185
er
n- .. 188
r- . 201
. 209
. 215
. 219
r . 229
. 235
240
245
253

Journal of Educational Psychology
1971, Vol. 62, No. 3, 179–184

INTERRELATIONS AMONG LEARNING AND PERFORMANCE TASKS IN DISADVANTAGED CHILDREN†

HAROLD W. STEVENSON,[1] ANNE MARIE WILLIAMS, AND EDGAR COLEMAN
University of Minnesota

Eight learning and eight performance tasks were administered to 50 4- and 5-year-olds attending day care centers. The results give little support to positions positing differences in the learning processes of middle- and lower-class children, for the intercorrelations were very similar to those previously obtained with middle-class children. Paired-associate learning and serial memory were significantly correlated, as were observational learning and paired-associate learning, serial memory, and category sorting. Paired-associate learning was significantly related to category sorting. The only performance task with a high frequency of significant correlations with the learning tasks was instructions. The frequency of significant correlations between instructions and the learning tasks was greater than that found among the learning tasks themselves.

This study was undertaken in an effort to provide additional information about individual differences in children's learning. It is, essentially, a replication of a previous study (Friedrichs, Hertz, Moynahan, Simpson, Arnold, Christy, Cooper, & Stevenson, 1971) of interrelations and correlates of learning at the preschool level. The major difference between the two studies lies in the children who were subjects. Whereas in the previous study the subjects attended a one-half day University laboratory preschool and were almost entirely from middle- and upper-middle-class families, the subjects in this study attended day care centers all day, 5 days a week, and were from lower socioeconomic groups.

The major question is whether the correlations among various measures of learning and performance differ according to the background of the children being tested. Some writers have proposed that differences would be found. Jensen (1969), for example, has proposed a continuum of learning abilities, ranging from what he calls Level I (associative learning) to Level II (cognitive learning). An example of a Level I task is paired associates, and of a Level II task, concept learning or problem solving. Jensen assumes that Level I abilities are necessary but not sufficient for Level II abilities, and, importantly, that the abilities are distributed differently among advantaged and disadvantaged children. Learning in disadvantaged children is characterized as being of the associative type, while advantaged children are assumed to learn in a more cognitive manner. If Jensen's proposals are valid, it would be expected that correlations between associative and cognitive learning would differ according to the group membership of the child.

In the previous study by Friedrichs et al. (1971) the median correlation among all learning tasks was .17, indicating a high degree of differentiation of learning abilities in preschool children. There were significant correlations, however, between some of the simple learning tasks and tasks dependent upon more complex cognitive processes. For example, paired-associate learning was significantly related to problem solving and to category sorting. These results do not depart from what might be predicted from a position such as Jensen's, for such correlations might be expected for middle-class children. More critical, however, is whether such correlations would be found for disadvantaged children.

Both simple associative and complex cognitive tasks are grouped under the general term, learning tasks. These tasks were adapted from those that have been used in laboratory studies and were selected on the basis of their possible interest to preschool children, their level of difficulty, and the

---

[1] Requests for reprints should be sent to Harold W. Stevenson, Institute of Child Development, University of Minnesota, Minneapolis, Minnesota 55455.

Copyright © 1971 by the American Psychological Association, Inc.

amount of time required for their completion. The following eight tasks were used: paired-associates, serial memory, oddity learning, concept formation, observational learning, incidental learning, problem solving, and category sorting.

The eight performance tasks were devised to tap attributes of potential relevance for understanding children's learning. They included social imitation, persistence, reactivity, impulsivity, following instructions, variability of behavior, attention, and level of aspiration. Each of the learning and performance tasks was administered individually to the subjects, 4- and 5-year-old preschool children.

METHOD

Subjects

The subjects were 23 girls and 27 boys attending day care centers in Minneapolis and St. Paul. The mean age of the subjects was 60.1 months ($SD =$ 8.6). There were 40 black, 7 white, 2 Indian, and 1 Spanish-American children. Although the majority of the children are classified as black, some of them were from interracial marriages. The children were in attendance at the day care centers 5 days a week, usually from 7:00 A.M. to 5:00 P.M.

Materials

The 16 learning and performance tasks are described in detail in Friedrichs et al. (1970). Their essential features are described in the following paragraphs.

Learning Tasks

1. *Paired associates.* Five pairs of common animals were depicted in line drawings. Each pair was presented and the child then was asked to select the proper associate for each stimulus element from the pool of response elements. The score was the number of correct response elements selected on each of three trials.

2. *Serial memory.* The subject was shown, one by one, five pictures of common objects. After all five pictures had been seen, the experimenter pointed to the location occupied by each picture and the subject was asked to recall its name. The score was the number of correct responses on three trials.

3. *Oddity learning.* On each of 36 trials the subject was presented three miniature objects; two were the same and one was different (e.g., two hearts and one football). Choices of the odd object revealed a plastic chip and the subject was told to find as many chips as he could. The score was the number of correct responses.

4. *Concept formation.* A board contained 30 rows of drawings of flowers, with three flowers in a row. Within each row the flowers were of a different shape and color. The subject was to insert a stylus in the flower he thought was correct in each row. A correctional procedure was used and the score was the number of first choices that were correct. The blue flower always was correct.

5. *Observational learning.* Two puzzles were constructed so that the five irregularly shaped pieces in each puzzle fit inside a 10 × 10 inch wooden frame in one specific spatial arrangement. The experimenter demonstrated how the pieces were to be inserted by removing each piece carefully and returning the pieces, in the same order, to their correct position. The score was the amount of time required to insert the pieces properly for each puzzle, thus a high score was indicative of poor performance. Two 2-minute trials were allowed with each puzzle.

6. *Incidental learning.* Three scenes involving children's activities corresponded with three simple stories. The stories directed the subject's attention to the scenes but did not make reference to specific items about which the subject later was questioned. After the pictures and stories were presented, the subject was tested for incidental learning. The subject was shown a second set of the scenes in which particular features were blank. The subject was asked to fill each blank spot by choosing one of four cards, only one of which depicted the correct alternative. The score was the total number of correct responses to the 16 questions that were asked.

7. *Problem solving.* In Problem Solving 1, the subject was shown a group of objects that could be inserted into a box in only one way. The objects were presented in disarray and the subject was asked to put them in the box so the lid would close. A maximum of 3 minutes was allowed and the score was the time required for solution.

In Problem Solving 2, an enclosed, L-shaped, clear-plastic maze contained a piece of "cheese" at one end and a plastic mouse at the other. An adjacent box contained a number of plastic rods. The child was encouraged to use the rods to move the cheese to the mouse. Solution required the subject to select a particular rod and to combine rods. Four points were given for successfully moving the cheese, two each for moving the cheese through the stem and the arm. If the subject succeeded in pushing the cheese through the stem with two unconnected rods, one point was given. There was a 5-minute time limit.

8. *Category sorting.* Three sortings of objects differing in size, brightness, form and/or height were required of each child. The procedure for sorting was demonstrated in a pretraining task with objects differing in two attributes (color and class) that were not represented in the test sortings. Nine objects were included in each test sorting, and each object differed on two attributes, such as size and brightness. A consistent sorting on the basis of one attribute yielded a score of 9, with a maximum score on the three sortings of 27. For imperfect sortings one point was assigned for each

TABLE 1

MEANS AND STANDARD DEVIATIONS

Learning task	*M*	*SD*	Performance task	*M*	*SD*
Paired associates	5.9	3.1	Social imitation	11.9	2.4
Serial memory	6.9	3.4	Persistence	245.2	51.7
Oddity learning	15.2	4.8	Reactivity	3.2	2.5
Concept formation	9.8	5.7	Impulsivity	14.5	16.9
Observational learning	5.5	2.8	Instructions	15.1	2.8
Incidental learning	5.8	1.8	Variability (a)	30.9	5.8
Problem Solving 1	2.6	.8	Variability (b)	2.8	1.3
Problem Solving 2	2.1	1.5	Attention	10.9	3.1
Category sorting	16.5	7.8	Aspiration (a)	2.2	.9
			Aspiration (b)	16.0	13.2

object sorted if there were at least two objects with the same attribute in each pile.

Performance Tasks

1. *Social imitation.* A Magic Treasure Box game was played in which the experimenter selected several objects (i.e. wand, scarf, hat) to be used in the game, made a series of verbal comments, and chose a particular treasure box. Each box contained a picture of an animal. The child, who followed the experimenter on each trial, was told only that it was now his turn. The score was the number of components of the experimenter's behavior imitated by the child, with a maximum score of 10.

2. *Persistence.* A square peg-board contained 289 small plastic pegs that could be pushed flush with the surface of the board. The child was told he could push them down one at a time and could stop when he wished. A maximum of 5 minutes was allowed and the score was the number of pegs depressed.

3. *Reactivity.* The subject operated a slide-changing button while seated in front of a rear-projection screen. He was allowed to project each of 35 pictures at his own rate. The score was the amount of time the subject spent looking at the pictures.

4. *Impulsivity.* A miniature truck containing "eggs" was to be taken to market. The experimenter stressed that the subject should go as slowly as possible so the eggs would not be broken. The score was the amount of time taken to push the truck along a straight 29-inch track.

5. *Following instructions.* A river, roads, landing strip, houses, bridge, and boat were represented on a large board. The subject was asked to carry out seven sets of instructions. The instructions involved such acts as taking a horse to the river and driving a truck to Farmer Jones' house. A total of 18 points was possible.

6. *Variability.* Thirty miniature wooden houses were mounted in five rows of six houses each on a large piece of plywood. The child was shown a car parked at a corner house and was told to drive the car to a friend's house which was centered on the opposite side of the board. The subject was given no instructions concerning the route to be taken. Two scores were used: (*a*) the number of intersections traversed in four trials, and (*b*) the number of different routes taken in the four trials.

7. *Attention.* The subject was seated before a rear-projection screen, given a response button, and told to press the button each time he saw a green ball. The green ball was projected so that the child would be familiar with it. Interspersed randomly among 65 slides were 15 identical slides depicting the greeen ball. Each slide appeared for 3 seconds. The score was the number of correct presses, with a maximum of 15. The subject was given a score on this task only if he pressed correctly at least once.

8. *Level of aspiration.* The subject was shown three towers differing in height and constructed of small, irregularly shaped wooden blocks. The subject was asked which tower he would like to build. After responding, he was given a box of 37 blocks and told to select as many as he wanted to build his tower. Two scores were assigned: (*a*) a score from 1 to 3 (shortest to tallest) for the tower chosen, and (*b*) the number of blocks selected by the subject for constructing his tower.

Procedure

Three day care centers were visited and all children of the appropriate ages were included in the study except for those who refused to cooperate. To obtain a sample of 50 children, 53 children were approached to participate in the study. Each child was tested individually for four sessions varying from 20 to 45 minutes. Four tasks were presented in each session. The order in which the tasks were presented was randomized across children. No prizes or other material rewards were provided.

The subjects were tested by one male graduate student and one female undergraduate student in psychology, each testing approximately equal numbers of girls and boys. Both experimenters were black.

RESULTS AND DISCUSSION

As may be seen from the standard deviations presented in Table 1, the children differed rather widely in their performance on each task. None of the learning tasks appeared to be particularly easy, nor was any of them so difficult that all of the children failed to attain some degree of success. Sex differences were assessed and in no task did the scores of girls and boys differ significantly.

Intercorrelations among Learning Tasks

As was found in the previous study (Friedrichs et al., 1971), the intercorrelations among the tasks were low. Only 6 of the 36 correlations were significant (see

TABLE 2

INTERCORRELATIONS AMONG LEARNING TASKS

Task	1	2	3	4	5	6	7	8
Paired associates (1)								
Serial memory (2)	.43**							
Oddity learning (3)	−.13	.16						
Concept formation (4)	−.17	−.06	−.07					
Observational learning (5)	−.31*	−.47**	−.03	.03				
Incidental learning (6)	−.02	.08	−.08	.26	−.21			
Problem Solving 1 (7)	−.09	.06	−.04	−.11	.13	.04		
Problem Solving 2 (8)	−.09	.06	.25	.18	−.05	.19	.23	
Category sorting (9)	.43**	.30*	−.11	−.09	−.35*	.19	−.03	−.19

* $p < .05$.
** $p < .01$.

Table 2). Scores on paired-associate learning and serial memory, the two associative learning tasks included in the group, were significantly correlated with each other. Observational learning was significantly related to paired associate learning, serial memory, and category sorting. Finally, category sorting, which presumably involved a form of concept formation, was significantly related to both paired associates and serial memory. These correlations are similar to those found in the earlier study, where 7 of the 36 correlations were significant. The only notable departure in the pattern of correlations involved Problem Solving 2. Whereas in the earlier study scores on this task were significantly related to those on paired-assocate learning, observational learning, and incidental learning, none of these correlations was significant in the present study.

Generally, the results do not give strength to the proposal that the learning process differs according to group membership. There is a great similarity in the results of the two studies, both in the correlations that were significant and in those that were not. Further, the significant correlations between associative learning tasks (paired associates and serial memory) and category sorting, a cognitive task, are not in accord with the assumption that disadvantaged children learn primarily in an associative manner. At least as far as the present results are applicable, and with the exception of Problem Solving 2, the learning process appears to be similar in the two groups of children. There is no ready explanation for the divergent results for Problem Solving 2. The failure to find significant correlations for this task cannot be attributed to a low level of performance or restricted variability, for the means and standard deviations found for this task in the present and previous study are remarkably similar ($M = 2.1$ versus 1.9; $\sigma = 1.5$ versus 1.6, respectively).

Correlations between Learning and Performance Tasks

The correlations between the learning and performance tasks are presented in Table 3. The task with the highest density of significant correlations was instructions, for which significant correlations were obtained with paired-associate learning, observational learning, incidental learning, Problem Solving 2, and category sorting. The correlations between scores on this task were more frequently significant than were the correlations among the learning tasks themselves. The pattern of correlations obtained for attention differed between the two studies. While attention was significantly correlated in the earlier study with oddity learning, observational learning, Problem Solving 2, and category sorting, the only significant correlation in this study was with serial memory.

Level of aspiration was negatively related to scores on observational learning, incidental learning, and Problem Solving 2; children who chose the largest number of blocks did the most poorly on these three

TABLE 3

CORRELATIONS BETWEEN LEARNING AND PERFORMANCE TASKS

Task	a	b	c	d	e	f	g	h	i
Social imitation	.03	.11	.27	.35*	−.10	.23	−.12	.18	−.16
Persistence	.17	.17	.10	−.08	−.14	.17	−.11	−.14	.20
Reactivity	.05	.29*	−.01	.13	−.27	−.05	.08	−.01	.39**
Impulsivity	.08	.08	.01	.07	−.03	−.12	−.14	.21	−.18
Instructions	.41**	.22	.10	.03	−.28*	.35*	.10	.32*	.31*
Variability (a)	.04	.09	.16	−.10	−.09	.10	−.05	.11	−.11
Variability (b)	.01	.06	−.06	−.19	−.03	.13	.08	.08	−.03
Attention	.13	.59**	.13	.08	−.08	−.01	−.01	−.06	.00
Level of aspiration (a)	−.18	−.15	−.08	−.03	.35*	.00	.10	−.17	−.13
Level of aspiration (b)	.06	.04	.08	−.15	.26	−.34*	−.03	−.36*	−.01

Note.—a = paired associates, b = serial memory, c = oddity learning, d = concept formation, e = observational learning, f = incidental learning, g = Problem Solving 1, h = Problem Solving 2, i = category sorting.

* $p < .05$.

** $p < .01$.

learning tasks. Scores on reactivity were positively related to those on serial memory and category sorting. Children who took the longer times to view the slides performed at a higher level on the two learning tasks. Thus, except for the instructions task, the performance tasks were of little assistance in interpreting the children's performance in the learning tasks.

Conclusions

In general, the correlations obtained in this and in the Friedrichs et al. (1971) study are notable for their similarity, rather than for their differences. From these data, it appears that there is little utility in positing differences between the two groups in the operation of associative and cognitive learning abilities. This is not to say that the performance of the two groups of children was identical. Generally, the disadvantaged children had more difficulty than the advantaged children in understanding what they were being asked to do, they required more preliminary instruction to familiarize them with the materials and procedures, and appeared to find some of the tasks, such as paired associates and category sorting, much more boring. In the paired-associate and serial memory tasks, many of the children did not know the names of the animals and objects, and had to be taught these names before the task could be administered. The procedure for category sorting had to be restated frequently before the children could comply with the instructions. Such difficulties contributed to the significantly lower scores of the disadvantaged, compared to the advantaged children on such tasks as serial memory ($t = 4.29$, $df = 96$, $p < .001$), oddity learning ($t = 3.15$, $df = 97$, $p < .001$), concept formation ($t = 3.55$, $df = 98$, $p < .001$), and category sorting ($t = 3.75$, $df = 94$, $p < .001$). It is interesting to note, however, that the differences were not significant for paired-associate learning, incidental learning, and problem solving. In observational learning, the disadvantaged children performed at a lower level than the advantaged children ($t = 2.26$, $df = 97$, $p < .05$). In view of this latter finding, it is interesting to note that the disadvantaged children received significantly higher scores in social imitation ($t = 7.12$, $df = 92$, $p < .001$).

Among the other performance tasks, there were significant differences between the two groups on persistence ($t = 2.93$, $df = 98$, $p < .01$), with the disadvantaged children remaining in the task a longer period of time, and on impulsivity ($t = 3.07$, $df = 96$, $p < .01$), with the disadvantaged children pushing the truck with greater speed. Differences could not be assessed for two tasks, instructions and variability. In the instructions task the instructions were read sentence by sentence in the present study, while in the previous study some of the more complex instructions were read as a series of sentences. Variability of performance was assessed in only four trials in this study, while five were used in the previous study. For the remaining performance tasks, reactivity, attention, and level of aspiration, no significant differences in performance were found.

It must be recognized that in comparisons of average levels of performance learning ability is confounded with prior knowledge. Given the deficiencies in disadvantaged preschool children's experience in formal learning situations and in their knowledge about common objects, perhaps a more revealing approach to understanding similarities and differences in the learning process of children from different social classes will be derived from a comparison of patterns of correlations among different learning tasks than from comparisons of differences in average levels of performance.

REFERENCES

Friedrichs, A. G., Hertz, T. W., Moynahan, E. D., Simpson, W. E., Arnold, M. R., Christy, M. D., Cooper, C. R., & Stevenson, H. W. Interrelations among learning and performance tasks at the preschool level. *Developmental Psychology,* 1971, **4,** 173–177.

Jensen, A. R. How much can we boost I.Q. and and scholastic achievement? *Harvard Educational Review,* **39,** No. 1, Winter 1969.

(Received February 19, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 3, 185–187

ACHIEVEMENT AS A FUNCTION OF THE PRESENCE OF PRIOR INFORMATION CONCERNING APTITUDE[1]†

ROBERT S. MEANS AND GLADYS H. MEANS[2]
Oklahoma State University

Seventy-two subjects enrolled in three sections of a class in adolescent psychology were told that they were being administered an aptitude test that had been used previously to help predict achievement in the course. The subjects were rank ordered by past grade point average. The median break was used to form two groups—high grade point average and low grade point average. The subjects within each group were randomly assigned to one of three treatments. Treatment 1 subjects were told that from the results of the aptitude test it would appear that they had high aptitude for mastery of the concepts taught during the course. Treatment 2 subjects were told that their test scores revealed little aptitude for the course. Treatment 3 subjects were given no information regarding their aptitude. A 2 × 3 analysis of variance design was employed to analyze student performances on the first teacher-made achievement test. Results indicated no significant differences from treatments, but significant differences were found favoring the high grade point average over low grade point average students. A significant interaction effect showed high grade point average students performing better with negative information whereas low grade point average students performed better with positive rather than negative information.

The influential role that knowledge of results on task performance may have on future related performance has often been cited in educational and psychological research. Locke (1967) states that the positive influence of knowledge of results on learning and performance is one of the best established findings in the research literature. However, few researchers have distinguished between the information or cueing function of knowledge of results and the motivational function. Locke hypothesizes that one possible means of isolating the motivational effect of knowledge of results is to give subjects knowledge that cannot be used for correcting errors on future tasks. To focus on the motivational effects of knowledge of results, it then becomes necessary to identify tasks that may not be used at all or may be used only in a limited way to correct errors on some future related performance task. Most of the reported research dealing with this dimension has been limited to simple computational, physical, or verbal tasks (Bilodeau, 1961; Chapanis, 1964). Attempts have also been made to assess the relationship that might exist between knowledge of results and level of aspiration (Feather, 1963), but this research too has mainly utilized simple tasks that are not academic in nature.

It was the intent of this study to investigate the possible motivational effect that knowledge of results of performance on some complex academic task might have on future performance in a related academic area. It would seem that academic aptitude tests whose scores are reported in general terms conform to the criteria of providing minimal cues for correcting future performance in a related area. Thus, it might be inferred that differences in information given students about their aptitude for future performance in some academic area would result in differing levels of motivation for achievement in that area. It would also seem likely that these different levels of motivation would be paralleled by differences in actual future achievement.

METHOD

This study was designed to test the following null hypotheses:

1. Differences in information given students con-

---

[1] This study was funded by the Oklahoma State University Research Foundation.

[2] Requests for reprints should be sent to Gladys H. Means, Department of Educational Psychology, Oklahoma State University, 2007 North Glenwood, Stillwater, Oklahoma 74074.

cerning their academic aptitude for adolescent psychology will not differentially influence levels of performance on a course content achievement test.

2. Differences in past academic achievement will not differentially influence levels of performance on a course content achievement test in adolescent psychology.

3. The results of a course content achievement test will not be significantly influenced by the interaction of past academic achievement and type of information given students concerning their academic aptitude for adolescent psychology.

Ninety-nine sophomore and junior students enrolled in three sections of an adolescent psychology course served as the initial pool of possible subjects for this study. All three sections were taught during the same semester and by one instructor. On the second meeting of each section all students enrolled and present were told that they were being administered an aptitude test that had been used in the past to predict achievement in the course. In reality the test was teacher made and had no previously demonstrated predictive validity. After testing was completed it was stressed to the students that even though the scores had been predictive of course success in the past, they would in no way be used to help determine final course grades. Students who enrolled late or who were absent on the day of testing were told that they would be given an opportunity to take the test at a later date. None of these later-tested students were used as part of the final sample studied.

The cumulative grade point average (GPA) through the previous semester was determined for each student in the initial testing group. These GPAs were taken from information in the school registrar's office. No student was aware that this information had been gathered. Students were rank ordered by GPA. The median-break method was employed to classify each student as either high or low in past GPA. Students who fell plus or minus five ranks from the median were thrown out of the possible pool of subjects. High GPA students were assembled in groups of three by consecutive rank. Each member of every three-member group was then assigned at random to one of three treatment groups. The same procedure was followed for low GPA students.

As part of the course requirements all students in the three sections were required to attend an individual conference with the instructor so that a term research topic could be assigned. These conferences were completed during the first 4 weeks of class. Students assigned to Treatment 1 were told in the conference that from the results of the aptitude test it would seem that they had high aptitude for mastery of the concepts taught during the course. Students who received Treatment 2 were given negative information; that is, they were told that their scores on the test reflected little aptitude for the course. Students assigned to Treatment 3 were given no information concerning the aptitude test during the conference. If a student from Treatment 3 or from those students whose GPAs had not been available asked specifically about his performance on the test, he was told that scoring of the test had not been completed but that he would be told his score when they were available. The test results were not discussed by the instructor in a regular class session.

A 2 × 3 multiple analysis of variance design was employed to analyze results of student performance on the first teacher-made multiple-choice course content test. This test was administered during the eighth week of class. After all class dropouts and students with no GPAs were eliminated from the sample population, 77 students remained. In order to equalize the number of persons in the various cells, it became necessary to randomly drop five people from four different cells. The final sample was composed of 72 students.

To insure that the data could meet the basic assumption of homogeneous variances, the $F_{\max}$ test was used. An F ratio of 2.52, a nonsignificant value, resulted.

RESULTS

Table 1 shows the mean number of correct items on the course content test for the various treatment groups.

The analysis of variance revealed that treatment alone had no significant influence on test scores, $F < 1.00$. Past GPA did have a significant influence. Those subjects with high GPAs attained significantly higher scores on the course content test than did those subjects with low GPAs, $F = 38.25$, $p < .001$. Treatment significantly interacted with past GPA to help shape the scores made on the course content test, $F = 5.56$, $p < .01$. An analysis of simple effects revealed that the significant interaction with GPA was between two of the stimulus levels —negative and positive information ($p < .01$). Specifically, high GPA students given negative information regarding their course aptitude attained significantly higher scores on the course content test than did those

TABLE 1

MEAN SCORES OF VARIOUS TREATMENT GROUPS ON COURSE CONTENT TEST

Group	Positive information	Negative information	No information	Combined
High GPA	42.00	47.75	45.92	45.22
Low GPA	39.58	35.00	37.50	37.36
Combined	40.79	41.37	41.71	41.29

Note.—N = 72 for the combined groups, 36 for high GPA, 36 for low GPA.

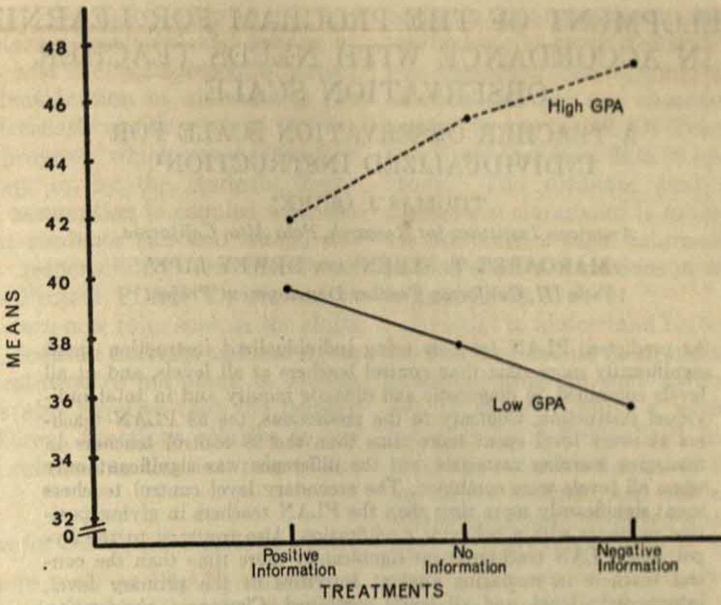


FIG. 1. Effect of aptitude information and achievement level on a course content test.

high GPA students given positive information regarding their aptitude ($p < .01$). Among low GPA students the opposite effect occurred. That is, those low GPA students given positive information regarding their course aptitude attained significantly higher scores on the course content test than did those low GPA students who were given negative information regarding their aptitude ($p < .05$). At both GPA levels students given no information regarding their course aptitude scores performed at a level near the combined cells' mean. This interactive effect is graphically presented in Figure 1.

The present study poses implications for the teacher, the counselor, and the researcher. From the evidence gathered it seems possible either to inhibit or to facilitate a student's performance by variation of the nature of information given him concerning his potential for performance in a given area. This investigation should be extended to other academic and nonacademic situations in order to assess the validity of the original findings. It would also be useful to investigate the possible relationships that may exist between type of prior reinforcement given and other organismic variables.

REFERENCES

BILODEAU, E., & BILODEAU, I. Motor-skills learning. *Annual Review of Psychology*, 1961, **12**, 243-280.

CHAPANIS, A. Knowledge of performance as incentive in repetitive, monotonous tasks. *Journal of Applied Psychology*, 1964, **48**, 263-267.

FEATHER, N. The relationship of expectation of success to reported probability, task structure, and achievement related motivation. *Journal of Abnormal Psychology*, 1963, **66**, 231-238.

LOCKE, E. Motivational effects of knowledge of results: Knowledge or goal-setting. *Journal of Applied Psychology*, 1967, **51**, 324-329.

(Received February 19, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 3, 188–200

DEVELOPMENT OF THE PROGRAM FOR LEARNING IN ACCORDANCE WITH NEEDS TEACHER OBSERVATION SCALE

A TEACHER OBSERVATION SCALE FOR INDIVIDUALIZED INSTRUCTION[1]

THOMAS J. QUIRK[2]
American Institutes for Research, Palo Alto, California
MARGARET T. STEEN[3] AND DEWEY LIPE[4]
Title III, California Teacher Development Project

As predicted, PLAN teachers using individualized instruction spent significantly more time than control teachers at all levels, and at all levels combined, in diagnostic and didactic inquiry and in total individual instruction. Contrary to the predictions, the 58 PLAN teachers at every level spent more time than the 28 control teachers in managing learning materials, but the difference was significant only when all levels were combined. The secondary level control teachers spent significantly more time than the PLAN teachers in giving positive messages within behavior modification. Also contrary to the hypotheses, PLAN teachers spent significantly more time than the control teachers in managing student activities at the primary level, intermediate level, and all levels combined. Classroom observation data were obtained in spring of 1969.

Programs for individualized instruction have recently put emphasis on instruction as a system to accommodate and expand individual differences in students. The beginnings of these systems can be seen in such programs as Individually Prescribed Instruction (Glaser, 1966), Program for Learning in Accordance with Needs (PLAN; Flanagan, 1967), the Primary Education Project (PEP), the NOVA schools, and ES70 Project through the Office of Education's Bureau of Research. These systems imply significant changes in the role of the teacher from those roles expected of a teacher in a traditional, self-contained classroom.

Program for Learning in Accordance with Needs

PLAN is a major program of individualized instruction which has two major goals: (*a*) to individualize the curriculum in the areas of social studies, language arts, mathematics, and science in Grades 1–12, and (*b*) to individualize the program of guidance and individual planning for each student within these curriculum areas.

The systems aspects of PLAN are designed to take care of those areas of instruction often assumed by the teacher but which could be more efficiently handled through technology or classroom organization. The scoring of achievement tests, the monitoring of the daily progress of students, the storage, retrieval and sorting of information, and the management of materials and identification of alternatives for curricular decisions are aspects of the instructional program which can be handled by technology or by the students themselves. This assumption is coupled with the concept that students can and should assume more responsibility for their own learning. In Project PLAN the student is required to learn new roles such as the skills of planning, decision making, and managing the implementation of his plans. A system such as this also requires that the teacher perform different functions relative to the system and relative to the new role of the student.

---

[1] The major support for this research was conducted under contract 68-05331 with the Office of Education, United States Department of Health, Education, and Welfare. School districts cooperating in this ESEA Title III project were: Archdiocese of San Francisco, Fremont Unified School District, San Carlos Elementary School District, San Jose Unified School District, Santa Clara Unified School District, and Union Elementary School District, all in the San Francisco Bay area. Funds were received and administered through the Fremont Unified School District. The results of this study were presented at the annual meeting of the American Psychological Association, Washington, D. C., September 1969. The authors gratefully acknowledge the assistance of Jan Bacigalupi, Helen Dell, Gary Ferney, Janet Henry, Christy Neushul, and William Stilwell who assisted in the definitions of the categories and Pris Johnston and William Brastow who assisted in the early phases of data analysis.

[2] Requests for reprints should be sent to Thomas J. Quirk who is now at the Developmental Research Division, Educational Testing Service, Princeton, New Jersey 08540.

[3] Now at Westinghouse Learning Corporation, Palo Alto, California.

[4] Now at American Institutes for Research, Palo Alto, California.

Implications for the Teacher Role in PLAN

In a system such as PLAN, the role of the teacher is to work within the system to insure the student's achievement of his unique set of objectives and to work with the student to assist him in acquiring the highest level of responsibility for managing his own learning. In the former role the teacher is the manager; in the latter role, a counselor. As a manager, the teacher assumes responsibility for monitoring the system, modifying the system, and managing those things which cannot be handled by the system. As a counselor, the teacher uses whatever strategies are necessary to decrease adult support for the student and to increase the student's responsibility for his own learning in those areas for which this emphasis is important for the individual student. The PLAN Teacher Observation Scale was designed to categorize teacher behaviors which can occur in both individualized instruction and in traditional instruction, but which are hypothesized to occur to a different degree within these two types of learning environments.

THE DEVELOPMENT OF THE PLAN TEACHER OBSERVATION SCALE

The PLAN Teacher Observation Scale was developed so that the behaviors of teachers in Project PLAN classrooms could be observed. This article relates to the applicability of the PLAN Teacher Observation Scale to two questions: (*a*) Do PLAN teachers as a group behave differently in the classroom than a group of non-PLAN teachers, and (*b*) can observers be reliably trained to use the PLAN Teacher Observation Scale to collect data in operating classrooms? The ultimate goal of observing teachers in classrooms is to understand better the teacher–pupil interaction and to be able to specify conditions in which learning is most likely to occur.

In order to understand better what teachers and students do in an individualized educational program such as Project PLAN, and in order to build a model of teacher behavior in individualized instruction, an objective record is needed of the actual behavior of teachers and students in classrooms using individualized instruction. Records of teacher behavior in the classroom require the use of an observational technique whereby information can be abstracted, recorded, and quantified (Remmers, 1963).

In order for the classroom observation instrument to be most useful, the observers should not be asked to record their personal perception of the "intent" of the teacher but rather to record the type of overt behavior which the teacher has demonstrated. Some degree of interpretation is inevitable, of course, but the training study attempted to keep the amount of interpretation to a minimum and the observers were asked to code how the behavior of the teacher looked to the pupils. The function of the observer is to abstract the behavior of the teacher into categories.

The presence of the observer in a classroom very probably has some effect on the way the teacher behaves in the classroom, but it is far better to have data on how teachers behave under observation than to have no data at all on how they behave in the classroom.

The emphasis in the PLAN Teacher Observation Scale is on the verbal behavior of the teacher. Nonverbal behaviors, such as gestures and facial expressions, are not recorded except in the two categories of Behavior Modification. The advantage of a classroom observation over an audiotape is that nonverbal behaviors can help to make the discriminatory task easier if they are included as a small part of the total observational procedure.

Several observation scales of teacher behavior have been developed for use in traditional classrooms. The Flander's Interaction Analyses technique (Amidon & Flanders, 1963), the Verbal Interaction Category System of Amidon and Hunter (1967), and Hough's (1967) Observational System for Instructional Analyses represent some of the better known instruments for the observation of teacher behavior. Medley and Mitzel (1958, 1959, 1963) developed an instrument named OScAR (Observation Schedule and Record) to quantify the behavior of beginning teachers so that the behavior could be correlated with a number of other variables. Flanders (1960) developed a sophisticated technique for observing the climate of the classroom which also preserves the sequence of teacher behavior. The observer who uses Flanders Interaction Analyses technique records a category number at the end of every 3-second period to represent the teacher's behavior during that interval of time.

All of the data reported here were collected by trained observers who observed teachers in their ongoing classrooms. The observer tallied the behavior of the teacher into one of 17 predefined categories in 5-second intervals. If more than one category of behavior occurred during the 5-second period, the observer was instructed to record only the latter behavior. A 5-second interval was chosen to allow the observer using the PLAN Teacher Observation Scale to record teacher behaviors along the same time constraints as a separate observer who simultaneously used the PLAN Student Observation Scale to record the behavior of the students in that classroom.

If the teacher walked about the room, the observer followed the teacher at a respectable distance in order to hear the comments of the teacher. The observer who recorded the behavior of the teacher did not record the behavior of the students. Another observer in the same classroom used the PLAN Student Observation Scale to record the behavior of the students (Lipe, Steen, & Quirk, 1969). The observer who used the Teacher Observation Scale was concerned only with teacher behavior and the behavior of the teacher in response to pupil behavior. The content of the pupil behavior therefore serves the same function as the nonverbal behavior of the teacher, and that is to aid the observer in classifying the teacher's behavior.

An important difference in the PLAN Teacher Observation Scale from other observation instruments is its distinction between the content and process[5] dimensions of teacher behavior and its emphasis on categories describing the systems management aspects of individualized instruction classrooms.

The observers trained to use the PLAN Teacher Observation Scale were permitted to change a category classification at a later time only under a single exception. The exception involved the change thereafter from Category 8 (tutoring) to Category 9 (lecturing) as soon as the teacher talked for a period of time greater than 60 seconds. For all other categories of teacher behavior, the observer was instructed to code each 5-second time period as a unit and then to forget that time period.

Definition of the PLAN Teacher Observation Scale Categories

The original PLAN Teacher Observation Scale was developed by having observers write down a detailed record of what teachers were doing in classes using individualized instruction. There were 255 separate categories in the first version of the PLAN Teacher Observation Scale but some behaviors occurred so infrequently that they were not worth recording as separate events. The scale was reduced to 12 categories, and finally expanded to 17 categories after preliminary reliability studies indicated that these

---

[5] "Content" refers to intra-Teaching–Learning Unit learning problems, that is, those associated with learning activities presented in a Teaching–Learning Unit, such as the understanding of the instructional objective and knowledge of the learning concepts. "Process," on the other hand, refers to extra-Teaching–Learning Unit topics, that is, the prerequisites to doing the learning activities, such as budgeting time, managing reinforcement activities, and locating instructional materials and equipment.

17 categories permitted the best description of events which would be expected to occur in PLAN classrooms, so that the categories would be clearly defined, distinct, reliable, and would reflect important differences among teacher activities in individualized instruction that can be easily discriminated by observers. The final 17 categories were tested originally by using videotapes of PLAN classes and scripted audiotapes.

OPERATIONAL DEFINITIONS OF TEACHER BEHAVIOR CATEGORIES

Teacher behaviors in PLAN classes are conceptualized as representing five major behavioral categories: (a) individual instruction, (*b*) group discussion, (*c*) behavior modification, (*d*) systems management and (*e*) other.[6]

Individual Instruction

This major category refers to a teacher's interaction with one student for the purposes of resolving a problem or extending concepts and interests. A teacher might question a student in order to determine the degree to which he has achieved a Teaching-Learning-Unit objective; he might ask the student to give examples that illustrate the objective, to explain the objective in his own words, or simply to tell what he learned. A teacher might then want to help the student relate the concept to other times or other societies.

Category 1: Diagnostic and didactic inquiry. This subcategory of individual instruction is represented by the teacher's attempts to clarify and probe into the source of problem areas. For diagnostic and didactic inquiry to occur, the inquiry would need to be followed up by identifying a learning problem and developing the problem toward some solution. (Examples: (*a*) "What did you do to review for the test?" (b) "Tell me in your own words what that objective means.")

Category 2: Decision facilitating. This subcategory of individual instruction refers to those behaviors of the teacher that invite participation of the student in the process of planning and decision making before a course of action is finalized. (Examples: (*a*) "What do you think you should do to prepare yourself better for the next group of objectives?" (*b*) "How many steps should we schedule you to complete each day in your social studies?")

Category 3: Solution giving. Teacher behaviors in this subcategory of individual instruction consist of the course of action for a student to take or direct explanation that resolves the student's problem. (Examples: (*a*) "Chartres is in France, not Belgium." (*b*) "You'd better work on that objective a little longer before you take the test.")

Category 4: Extending concepts and interests. This subcategory occurs when the teacher takes an opportunity during individual instruction to relate the student's current learning activities to other areas of the student's knowledge, interest, and experience. (Examples: (*a*) Asking the student to relate the concept to other times and societies or to present day society; (*b*) giving the student several examples and asking him to give a generalization.)

Category 5: Silent attending. This subcategory consists of the teacher's nontalking behavior during individual instruction. (Examples: (*a*) Pausing while a student thinks through his answer; (*b*) listening to a student's response to the teacher's question.)

Group Discussion

This major category refers to a teacher's interaction with two or more students during a group interaction. Functions of the teacher in group discussion include modeling the group discussion leader role and teaching students how to lead a group discussion so that eventually students can carry on an effective group discussion without the teacher's presence. In a group discussion the teacher may also function as a tutor of the group or he may lecture.

Category 6: Modeling the discussion leader role. This subcategory of group discussion refers to those specific teacher statements and activities that indicate to group participants which group leader behaviors they are to imitate. (Examples: (*a*) "Did you notice that your discussion leader asked Susan to elaborate her point so that you all would understand more clearly what she meant? Remember that so you can do it when you lead a group discussion sometime." (*b*) "You'll notice I didn't call on Johnny even though his hand was up. I didn't call on him because he had already talked more than other members of the group.")

---

[6] Since this article represents the first formal publication of data collected using the categories of the PLAN Teacher Observation Scale, brief definitions and examples of the categories are given here. For a more complete description, see the *Manual for Training Classroom Observers to use SOS and TOS* (1969).

Category 7: Leading group discussion. The discussion leader's role can include defining the topic, stating the criteria for evaluating discussion, pointing out unsubstantiated statements, calling on nonparticipants, limiting overparticipation, asking for clarification of a statement, and summarizing major points. (Examples: (*a*) Reminding participants that opinions are to be supported by facts; (*b*) asking one of the silent students what his opinion is.)

Category 8: Tutoring (discussion). This subcategory of group discussion refers to a teacher's conducting a group tutoring session in which problems are identified and resolved through a question and answer procedure. Tutoring is distinguished from lecturing (Category 9) in that tutoring involves active participation by students whereas lecturing requires only quiet listening on the part of the group members. (Examples: (*a*) Extending concepts through a group discussion; (*b*) asking for a student volunteer to describe in his own words the objective being discussed.)

Category 9: Providing content (lecturing). This subcategory of group discussion refers to the teacher's talking for an extended period of time. If the teacher interjects into the discussion an extended explanation or other information which takes longer than about 60 seconds, it would be thereafter categorized as lecturing. (Examples: (*a*) Showing students how to work a long division problem; (*b*) describing the plot of a story.)

Category 10: Silent attending. This subcategory consists of the teacher's nontalking behavior during group discussion. (Examples: (*a*) Waiting for a student to volunteer a comment; (*b*) listening to one student challenge another.)

Behavior Modification

This major category occurs at any time when the teacher communicates to a student a positive or negative verbal or nonverbal message for the purpose of increasing or descreasing the probability that the student's behavior will occur again.

Category 11: Giving positive verbal or nonverbal message. This subcategory refers to teacher approval of desirable behavior by praise or gestures. (Examples: (*a*) "I'm so pleased that you found that book without asking my help"; (*b*) patting a student on the shoulder when he is doing a desirable activity.)

Category 12: Giving negative verbal or nonverbal message. This subcategory refers to teacher disapproval of undesirable behavior. (Examples: (*a*) Verbally criticizing some students for scuffling; (*b*) staring and frowning at students who are joking and laughing.)

Systems Management

This major category consists of those teacher behaviors that regulate the flow of information and learning tools in order to facilitate each student's learning activities. It concerns classroom logistics in that it functions to keep students progressing from one activity to the next.

Category 13: Managing records and computer materials. This management function includes the teacher's working on performance records, computer cards, and resolving student problems with the mechanical and procedural aspects of the use of computer materials. (Examples: (*a*) Checking test cards for marking errors; (*b*) asking a student what unit he is working on so that the teacher can mark this in his record book.)

Category 14: Managing learning materials and equipment. Teacher behaviors in this management subcategory include distributing Teaching–Learning Units, tests, materials, and equipment, repairing equipment, and logistics problems about materials. (Examples : (*a*) Assembling materials for a science experiment; (*b*) placing new learning units in the file.)

Category 15: Managing student activities. This management subcategory concerns brief interactions with students which usually concern the distribution and placement of students throughout the classroom. (Examples: (*a*) "All right, children, get your learning units and start to work." (*b*) "Some of you in the back of the room, go over to the language arts center to work.")

Category 16: Observing, listening, or walking. This management subcategory refers to a teacher's monitoring student activity. (Examples: (*a*) Walking around the classroom observing students at work; (*b*) listening to a student's question (but different from Category 5 or 10).

Other

This major category is a catch-all for behaviors unrelated to instruction.

Category 17: Activities unrelated to instruction. Teacher behaviors belong in this subcategory if they cannot properly be placed in any of the other categories. These behaviors include silence or confusion that sometimes occurs (e.g., just before recess or the end of the period) and socializing by the teacher (e.g., telling a girl how nice her hair looks).

Selecting and Training Observers

The training program for classroom observers lasted 7 days. The first half of the training program was devoted to the training of the observers on the PLAN Student Observation Scale, and the results of this training are reported elsewhere (Lipe et al., 1969). The second half of the training program was devoted to the PLAN Teacher Observation Scale and involved written materials which presented background information on Project PLAN, practice exercises in categorizing written examples of teacher behavior, discussion of the operational definitions of the 17 categories, visits to PLAN classrooms with the trainer who pointed out examples of the behaviors relevant to the 17 categories, practice coding in ongoing classrooms with another observer and subsequent computation of the resulting Scott's π by the observers for these practice trials, and discussion of coding problems with the trainer. A complete summary of the training program is given in the *Manual for Training Observers to Use SOS and TOS* (1969).

A group of four women (Group 1) was trained during the middle of March 1969 to use the teacher observation form. The training procedure was replicated with another group of four women (Group 2) during the last 2 weeks of May 1969. The first group of four women ranged in age from 23 to 35. Their educational experience ranged from 1 year of college to 1½ years of graduate school. Two of the group had had no teaching experience, the third had 3 years teaching experience, and the fourth 5 years of teaching experience.

The procedure upon entering the classroom was to start their stop watches at exactly the same time. Then, staying close enough to the teacher to hear what was said, the observer recorded the teacher's behavior each time the stop watch sweep second hand reached a 5-second point. The process soon developed into a smooth rhythm of observing for 3 seconds, deciding during the fourth second which category represented the behavior, and recording during the fifth second. If a teacher switched behaviors during the 3 seconds (e.g., if he changed from talking to attending to a student), the latter behavior of the interval was recorded. If a teacher performed two behaviors simultaneously, both behaviors were recorded.

The reliability study was designed to include eight 15-minute observations each at the primary, intermediate, and secondary levels. Observers were randomly assigned to the letters A through D, and then they observed classes according to the predetermined pairs of observers indicated in Table 1. Each team observed two Project PLAN classes and two traditional classes at one of the levels.

During the first day, eight primary and eight intermediate level observations were made. The reliabilities were computed the same evening, and it was found that Observer A with both her partners for the day achieved only two out of eight reliabilities above .85 and one reliability was .00 (see Table 1). The next morning, before beginning observations in the secondary classes,

TABLE 1

OBSERVER RELIABILITY

Grade level observed	Group 1 Scott's π for				Group 2 Scott's π for			
	Project PLAN classes		Traditional classes		Project PLAN classes		Traditional classes	
Primary (Grades 1 and 2)								
Observers A and B	.81	.70	.72	.72	.96	.89	.92	.91
Observers C and D	.85	.82	.85	.97	.88	.79	.87	.95
Intermediate (Grades 4, 5, and 6)								
Observers A and C	.61	.00	.94	.88	.88	.90	.89	.91
Observers B and D	.84	.89	.98	.85	.88	.92	.91	.71
Secondary (Grades 9 and 10)								
Observers A and D	.90	.91	.92	.95	.90	.95	.97	.95
Observers B and C	.67	.83	.87	.95	.81	.85	.93	.85
Observers A and B[a]	.91		.87					
Observers A and C[a]	.95		.95					

Note.—A new group of observers was trained to observe PLAN and control classrooms during fall of 1969. Both Scott's π and Spearman rank correlation coefficients were computed on the nine classroom observations obtained to check observer reliability. Scott's π ranged from .34 to .97 with a median π of .88. Spearman rank correlation coefficients on the same data ranged from .79 to 1.00 with a median coefficient of .99.

[a] Reliabilities of the extended practice sessions for Observer A of Group 1.

the observers and the trainer examined the discrepancies in observation with Observer A. It was discovered that Observer A had a misconception regarding the distinction between tutoring and lecturing during group discussion and had not adequately understood four of the other categories. It was decided to pair Observer A again with Observers B and C for four additional observations at the end of the reliability study. These are also reported in Table 1. All reliabilities of the extended practice for Observer A are above .85. Eighteen out of 28 total reliabilities are above .85. For the actual data collection, Observer A was not used for the PLAN Teacher Observation Scale.

The second group of female observers ranged in age from 22 to 56. All four had earned a bachelor's degree and only one had had any teaching experience. That teaching experience was 1 year in length.

The design of the reliability study with Group 2 was the same as that of Group 1. The results are given in Table 1 and only three of the coefficients are less than .85.

DATA COLLECTION

The training of teacher observers was completed March 25, 1969, but actual observation of teachers for data gathering purposes was not begun until the week after Easter recess ended. Observers C and D of Group 1 collected all the Teacher Observation Scale data reported in this article. Because of the 2-week lapse without practice, the two observers who were to obtain the teacher behavioral data rechecked their reliability before beginning their observations for data. They observed six classes, four Project PLAN and two traditional. Their reliability coefficients ranged from .89 to .97.

The observations were organized so that all 66 Project PLAN teachers in the 14 San Francisco Bay Area schools participating in Project PLAN and 32 randomly selected non-PLAN control teachers would be observed for three separate 20-minute observational sessions for a total of 1 hour's observation of each of 98 teachers. The distribution of Project PLAN and control teachers across elementary grade levels and secondary subject areas is shown in Table 2. Observations were made by two teams of observers. One member of each team observed and recorded teacher behavior while the other member concurrently observed and recorded student behavior. The results of student observations are reported else-

TABLE 2
DISTRIBUTION OF PROJECT PLAN AND CONTROL TEACHERS

Type of teacher	Number of elementary teachers by grade level				Number of secondary teachers by subject taught (Grades 9–11)				Total
	1	2	5	6	English	Social Studies	Math	Science	
Project PLAN teachers									
Number of teachers	12	12	12	12	5	5	4	4	66
Number observed for 1 hour	10	11	10	11	4	5	4	3	58
Control teachers									
Number randomly selected	5	5	5	5	3	3	3	3	32
Number observed for 1 hour	4	4	5	5	2	3	2	3	28

where in a separate report (Lipe et al., 1969).

The control teachers were randomly selected from the same schools as the Project PLAN teachers. At the elementary level five schools were randomly selected separately at each grade level. No school was asked to contribute more than three control teachers, however, and as soon as a school's name had been drawn three times it was dropped from the list. From each school that was selected, the name of a non-PLAN teacher at that grade level was randomly selected. Two elementary schools in one district had no non-PLAN teachers at the desired grade levels so the control teachers were drawn from a third school in that district, a school not participating in Project PLAN. Three of the control teachers selected were rejected by their respective principals because, in each case, the principal believed the teacher would be too uncomfortable with observers in the room. To replace each of the three rejected teachers, another teacher was randomly selected from the remaining teachers at the same school as the rejected teacher. Three control teachers were randomly selected from each of the four secondary schools participating in Project PLAN. None of the secondary teachers were rejected.

The principal or another administrator in each school notified both the Project PLAN teachers and the control teachers that observers would visit their classes on three separate occasions over the period of the next 4 to 6 weeks. He explained that their visits would be unannounced, that the data would not be reported to any school official or be seen by anyone other than the observers and the research team, and that the teachers should proceed with their normal activities whether or not the observers were present. The control teachers' participation was optional in all schools and only one refused to participate.

The two observer teams each scheduled their own visits to schools. One team began on one side of San Francisco Bay while the other team began on the other side, and they worked toward a middle point at the end of the Bay. When they met at the middle, most of the teachers had been observed once and the observers again checked their reliability. Their reliabilities were .96 in a secondary Project PLAN class, .99 in an intermediate non-PLAN class, and .97 in a primary Project PLAN class.

For various reasons individual teachers, and in one case a whole school, could not be visited. One Project PLAN teacher contracted pneumonia and only one observation was made in her class. Another teacher, a control teacher, politely refused the observers entrance to her room. One school was not visited at all because parent visitation week, standardized testing, and other special activities occupied too large a portion of the observation period to permit completion of 1 hour's observation of the teachers over three visits. The distribution of Project PLAN and control teachers that were observed for 1 hour is shown in Table 2.[7]

---

[7] Due to logistical difficulties in scheduling observers, one Grade 5 control teacher was observed

Each observer kept a record of the day of the week, date, and time of day of each observation. After most of the teachers had been observed once, that is, at the time the observers rechecked their reliability, they exchanged records of visitation times. Each team tried to plan the second visit on a different day of the week from that of the first visit and, at the elementary grade levels, on a different hour of the day. After most of the teachers had been observed twice, once by each observation team, the decisions regarding who would observe which schools for the third and last visit were worked out between the observer teams. In order to obtain three observations on as many teachers as possible, observers during the final phase were permitted to observe two intermediate-level PLAN teachers and two secondary-level control teachers twice on the same day, three of these teachers for two consecutive 20-minute observations. All other teachers were observed on three different days over the period from April 11 through May 29, 1969.

If a control class was taking a test or if a movie was being shown to the class at the time the observers visited, the observers simply left and returned at another time. There were a few occasions when PLAN classes were taking group standardized tests at the time of the observers' visits and observation had to be postponed. No substitute teachers were observed and the observer did not notify the teachers in advance of their visit. The observers did phone the principal's office on the day before visiting the classroom to notify the principal of their visit and to check on the administrative schedule for the following day. The hypotheses which were to be tested were not discussed with the observers during the training or during the data collection.

HYPOTHESES

A 3-day training conference was held for all western PLAN teachers in late August 1968 which included a film and completed practice exercises in diagnosing learning problems in students, training materials for decision facilitating within individual instruction, programmed booklets on student-managed behavior, training materials on classroom organization and the function and operation of the computer services within Project PLAN.

The emphasis on the preservice August training program and throughout the inservice training programs during the school year was to train the teachers to train the students to assume responsibility for their learning by using instructional materials and equipment independent of teacher supervision so that the PLAN teachers could be relieved of this administrative duty.

The August 1968 conference included an additional set of programmed materials which dealt with student managed behavior. These materials were followed up by a 2-hour in-service meeting in January 1969 between a field consultant and the teacher to discuss classroom activities to date which related to these training materials.

The August 1968 training conference also included materials on testing and a videotape model on the use of planning strategies and practice by the teacher with a student in the use of these strategies. One-half of the PLAN teachers also viewed a videotaped model of tutoring strategies and practiced these techniques with a student while videotaping their performance. Eight of the 58 PLAN teachers reported here did not attend the August 1968 conference, but in every case, except for the videotape treatments, the field consultants carried the printed materials to the teachers in their classrooms early in the fall.

Since most of the content necessary to achieve the instructional objectives is presented in Project PLAN in the Teaching–Learning Units, and since each student receives a program of studies through which he proceeds at his own best pace, lecturing to the entire class is an inappropriate behavior in many instances within a PLAN classroom. For a further discussion of the differences between a PLAN classroom and a traditional classroom, see Flanagan (1967; 1968) and Quirk (1970).

Since the present version of the PLAN

---

for only 55 minutes instead of 60 minutes, but since the teachers were ranked by the percent of time spent in each category, this logistical problem should not have been a serious one.

Teacher Observation Scale did not permit analyses of the data into separate categories for large-group discussion and small-group discussion, no specific hypotheses are presented regarding tutoring within group discussion or in leading group discussion.

In order to compare PLAN and control teachers on classroom observations on behaviors relevant to the management of Project PLAN and to the August 1968 training conference, the following hypotheses are presented:

Hypothesis 1. PLAN teachers will spend significantly more time than control teachers in diagnostic and didactic inquiry, in decision facilitating, in giving positive messages within behavior modification, and in total individual instruction.

TABLE 3

MEAN PERCENT OF TIME SPENT IN EACH ACTIVITY BY PLAN AND CONTROL TEACHERS

Category	Primary level				Intermediate level				Secondary level			All levels combined			
	PLAN ($n = 21$)	Control ($n = 8$)	Mann-Whitney		PLAN ($n = 21$)	Control ($n = 10$)	Mann-Whitney		PLAN ($n = 16$)	Control ($n = 10$)	Mann-Whitney *U*	PLAN ($n = 58$)	Control ($n = 28$)	Mann-Whitney	
			U	*z*			*U*	*z*						*U*	*z*
Individual instruction															
1	19.8	5.08	11	3.58**	23.94	9.81	22	3.5**	19.26	4.55	18**	21.16	6.57	246	5.22**
2	.08	.00	76	ties large	.15	.00	85	ties large	.32	.23	78 ties large	.17	.08	704	ties large
3	.72	.52			2.35	1.35			1.45	.44		1.51	.79		
4	.18	.00			.31	.00			.28	.05		.25	.01		
5	14.71	1.89			9.43	3.38			6.29	1.55		10.46	2.30		
Group discussion															
6	.00	.00			.00	.00			.00	.00		.00	.00		
7	.00	.06			.00	.05			.15	.24		.04	.12		
8	7.42	24.43			6.36	23.92			7.24	33.93		6.99	27.65		
9	.37	4.13	47	ties large	.05	3.80	34	ties large	1.26	9.59	48 ties large	.50	5.97	392	ties large
10	5.08	21.51			3.85	19.45			5.42	14.28		4.73	18.19		
Behavior modification															
11	1.97	1.38	61.5	1.11	.86	.44	67	1.61	.19	.58	35**	1.08	.76	658	1.43
12	1.00	.64	67	.83	.80	1.21	100	.21	.59	.65	73	.82	.85	785.5	6.25
Systems management															
13	.54	.12			3.55	.30			4.48	1.16		2.72	.56		
14	9.84	8.76	65.5	.90	8.58	6.36	86	.80	14.94	8.42	54	10.82	7.79	624	1.74*
15	24.45	18.09	41.5	2.08*	24.68	14.96	25	3.38**	17.97	12.66	49	22.72	15.03	381.5	3.97**
16	13.10	12.53			14.49	14.61			18.59	10.08		15.14	12.39		
Other															
17	.61	.79			.51	.27			1.49	1.51		.82	.87		
Total individual instruction (1+2+3+4+5)	35.55	7.50	7.5	3.74**	36.21	14.55	24	3.42**	27.62	6.84	19**	33.57	9.77	162.5	6**
Total group discussion (6+7+8+9+10)	13.08	50.14			10.27	47.24			14.07	58.06		12.35	51.95		
Total behavior modification (11+12)	2.79	2.03			1.67	1.66			.79	1.23		1.83	1.61		
Total system management (13+14+15+16)	47.93	39.52			51.31	36.25			55.99	32.33		51.41	35.78		

* $p < .05$.

** $p < .01$.

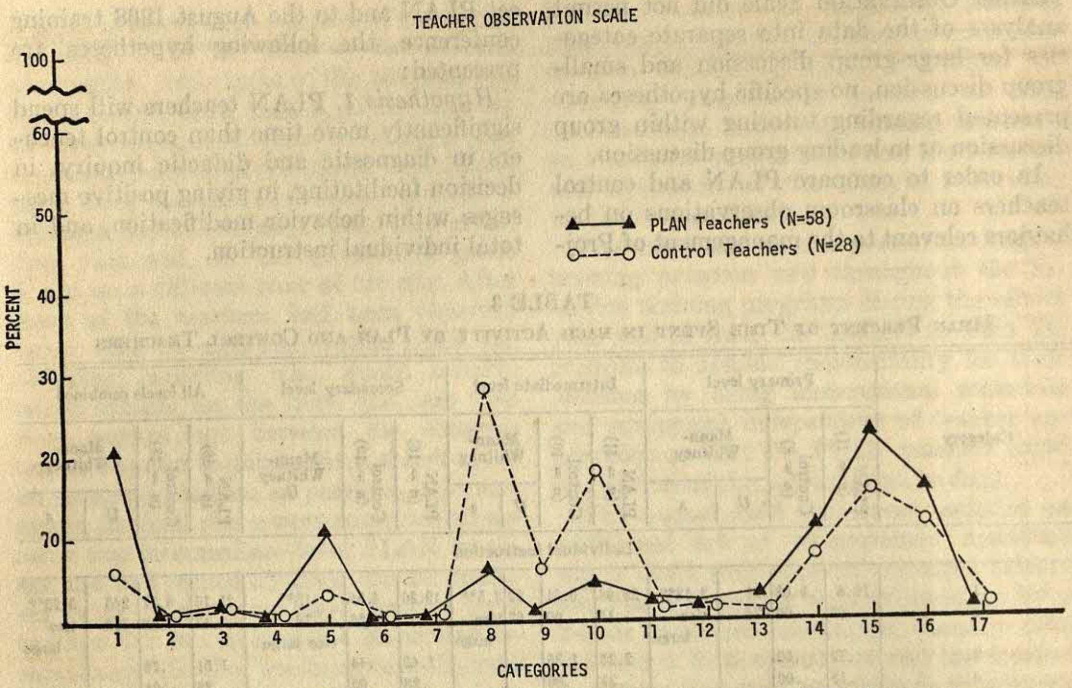


FIG. 1. Comparison of PLAN and control teachers when primary, intermediate, and secondary levels are combined.

Hypothesis 2. Control teachers will spend significantly more time than PLAN teachers in providing content within group discussion, in giving negative messages within behavior modification, in managing learning materials, and in managing student activities.

TESTS OF THE HYPOTHESES AND DISCUSSION OF RESULTS

All of the hypotheses were tested by means of a one-tailed, Mann-Whitney *U* Test (Hays, 1963; Siegel, 1956). In each case the PLAN and control teachers were ranked in terms of the percent of time which each teacher spent in the activity in question and the tabled comparisons of PLAN and control teachers report the mean percent of time spent in each activity by the two groups of teachers. The data for the classroom observations of the teachers at the primary level, intermediate level, and combined primary, intermediate, and secondary level teachers were corrected for ties. The data for the secondary level teachers were not corrected for ties because $N <$ 20 for the larger group of teachers, and therefore the significance level for the secondary level teachers is a conservative estimate of the differences between the two groups of teachers.

In order to increase the number of teachers in each group, the Level 1 and Level 2 teachers were combined into a single group of primary level teachers. Similarly, the Level 5 and Level 6 teachers were combined to form the groups of intermediate level teachers, and the Level 9 and Level 10 teachers were combined to form the groups of secondary level teachers.

Summary of PLAN and Control Teacher Comparisons

The results of the classroom observations for teachers at the primary level, intermediate level, secondary level, and all levels combined are summarized in Table 3. The data for all levels combined are also summarized in Figure 1 and Figure 2. As predicted, PLAN teachers spent significantly more time than control teachers at primary level, intermediate level, secondary level, and all levels combined in diagnostic and didactic inquiry (Category 1) and in

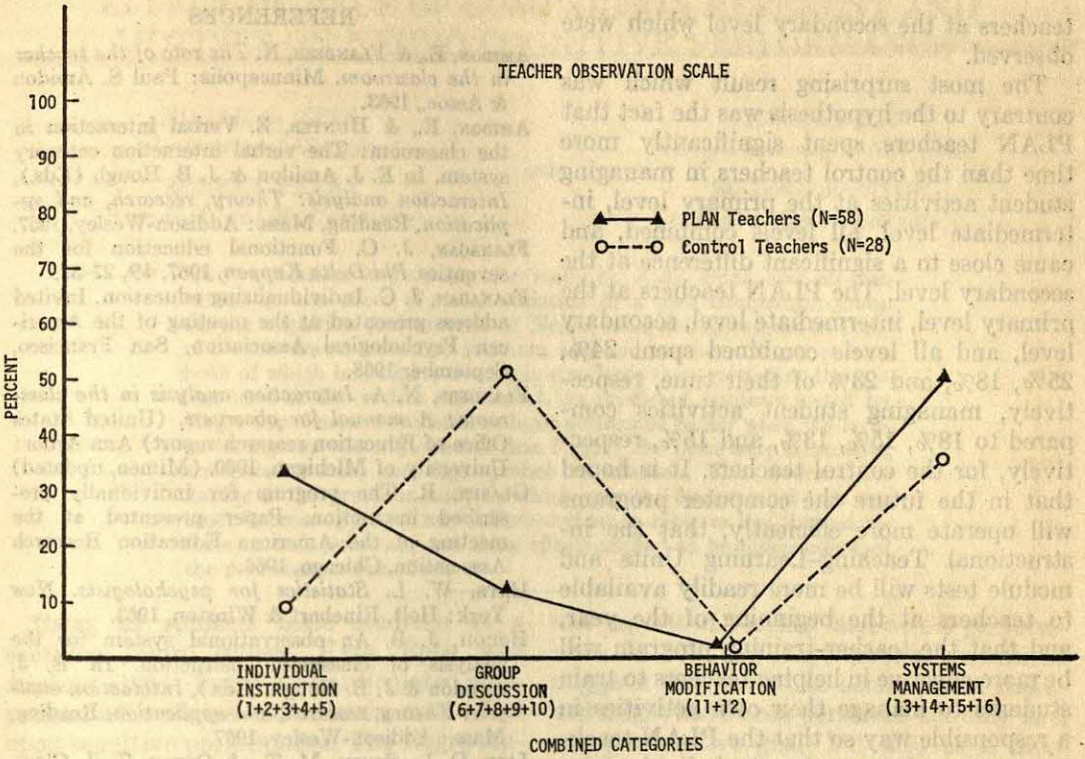


FIG. 2. Comparison of PLAN and control teachers on combined categories when primary, intermediate, and secondary levels are combined.

total individual instruction (Categories 1+2+3+4+5). The PLAN teachers at these levels spent 20%, 24%, 19%, and 21%, respectively, in diagnostic and didactic inquiry compared to 5%, 10%, 5%, and 7%, respectively, for the control teachers. PLAN teachers also spent 36%, 36%, 28%, and 34%, respectively, in total individual instruction compared to 8%, 15%, 7%, and 10%, respectively, for the control teachers at these levels.

There were no significant differences between the PLAN and control teachers at any level, or when the levels were combined, in decision facilitating, providing content within group discussion, or in giving negative messages within behavior modification. In the cases of decision facilitating and in providing content within group discussion, the differences were in the predicted direction, but the large number of ties at the identical percent at each level and at all levels combined provided an overestimate of the degree of difference between the two groups of teachers.

Contrary to the predictions, the PLAN teachers at every level spent more time than the control teachers in managing learning materials, but the difference was significant only when all levels were combined. When all levels were combined, PLAN teachers spent 11% of their time managing learning materials compared to 8% for the control teachers.

It was also surprising to find that, at the secondary level, control teachers spent significantly more time than the PLAN teachers in giving positive messages within behavior modification, but there were no significant differences between the PLAN and control teachers in this activity at the primary level, intermediate level, and all levels combined. Both groups of teachers, however, spent less than 2% of their time at each level and at all levels combined in this activity. This finding could be either a byproduct of the administrative schedule requirements at secondary level or a behavior typical of the particular group of PLAN teachers at the secondary level which were observed.

The most surprising result which was contrary to the hypothesis was the fact that PLAN teachers spent significantly more time than the control teachers in managing student activities at the primary level, intermediate level, all levels combined, and came close to a significant difference at the secondary level. The PLAN teachers at the primary level, intermediate level, secondary level, and all levels combined spent 24%, 25%, 18%, and 23% of their time, respectively, managing student activities compared to 18%, 15%, 13%, and 15%, respectively, for the control teachers. It is hoped that in the future the computer programs will operate more efficiently, that the instructional Teaching–Learning Units and module tests will be more readily available to teachers at the beginning of the year, and that the teacher-training program will be more effective in helping teachers to train students to manage their own activities in a responsible way so that the PLAN teachers can spend more time in individual instruction and in group discussion. The fact that there are not any classroom observations from the fall of 1968 prevents us from reporting whether the PLAN teachers decreased or increased the amount of time they spent in managing student activities from the fall to the spring, but classroom observation during the 1969–70 school year should help to clarify the trend in this activity.

During the 1969–70 school year the PLAN Teacher Observation Scale was expanded to include categories for large-group discussion (i.e., discussion by the teacher with more than one-half of the class) and categories under a major heading for small-group discussion (i.e., disussion by the teacher with at least two students but less than one-half of the students in the class) instead of the single major category for group discussion in the present version of the PLAN Teacher Observation Scale. In this way PLAN and control teachers can be compared better in the amount of time they spend in different activities within different types of group discussion instead of just one type of group discussion.

REFERENCES

Amidon, E., & Flanders, N. *The role of the teacher in the classroom.* Minneapolis: Paul S. Amidon & Assoc., 1963.

Amidon, E., & Hunter, E. Verbal interaction in the classroom: The verbal interaction category system. In E. J. Amidon & J. B. Hough (Eds.), *Interaction analysis: Theory, research, and application.* Reading, Mass.: Addison-Wesley, 1967.

Flanagan, J. C. Functional education for the seventies. *Phi Delta Kappan,* 1967, **49,** 27–32.

Flanagan, J. C. Individualizing education. Invited address presented at the meeting of the American Psychological Association, San Francisco, September 1968.

Flanders, N. A. *Interaction analysis in the classroom: A manual for observers.* (United States Office of Education research report) Ann Arbor: University of Michigan, 1960. (Mimeo, updated)

Glaser, R. The program for individually prescribed instruction. Paper presented at the meeting of the American Education Research Association, Chicago, 1966.

Hays, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

Hough, J. B. An observational system for the analysis of classroom instruction. In E. J. Amidon & J. B. Hough (Eds.), *Interaction analysis: Theory, research, and application.* Reading, Mass.: Addison-Wesley, 1967.

Lipe, D. L., Steen, M. T., & Quirk, T. J. Classroom behavior of PLAN students compared with control students. Paper read at the meeting of the American Psychological Association, Washington, D. C., September 1969.

Manual for training classroom observers to use SOS & TOS. (Title III Teacher Development Project, Fremont Unified School District. Contract 68-05331 with the Office of Education, United States Department of Health, Education, and Welfare) Fremont, Calif.: Fremont Unified School District, 1969.

Medley, D. M., & Mitzel, H. E. A technique for measuring classroom behavior. *Journal of Educational Psychology,* 1958, **49,** 86–92.

Medley, D. M., & Mitzel, H. E. Some behavioral correlates of teacher effectiveness. *Journal of Educational Psychology,* 1959, **50,** 239–246.

Medley, D. M., & Mitzel, H. E. Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.

Quirk, T. J. The student in Project PLAN: A functioning program of individualized education. *Elementary School Journal,* 1970, **71,** 42–54.

Remmers, H. H. Rating methods in research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

(Received February 20, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 3, 201–208

EFFECT OF RACE OF COMPARISON REFERENT AND MOTIVES ON NEGRO COGNITIVE PERFORMANCE[1]

EDGAR G. EPPS
University of Chicago

AUBREY PERRY
Florida A. and M. University

AND

IRWIN KATZ[2]
City University of New York

EUGENE RUNYON
Central State University

Two replications were done of a study by Katz, Epps, and Axelson (1964) on the effect of racial variation of a peer comparison group on the cognitive performance of Negro male college students. Samples were drawn from one southern and one northern Negro college, both of which had higher academic standards than existed at the college used in the earlier experiment. As predicted, subjects tested by Negro experimenters at both colleges performed better when the comparison group was white rather than Negro. The tasks were arithmetic, scrambled words, and digit-symbol substitution. This reversed the finding of the previous study, in which only the digit-symbol task was used. Success-approach and failure-avoidance motives had limited effects on task behavior, and these effects tended to be independent of the peer comparison effect.

Using as subjects male students at a southern Negro college, Katz, Epps, and Axelson (1964) found that anticipated comparison with whites had a detrimental effect upon cognitive performance. The Negro students were less proficient at a digit-symbol substitution task, administered by a Negro experimenter, when told their scores would be compared with white norms than they were when anticipating comparison with Negro norms. The present research was designed to investigate further the effect of intraracial and interracial competition on Negroes' cognitive performance.

Subjects in the Katz, Epps, and Axelson study reported that they "cared" more about doing well in the white norm condition than in the Negro norm condition, a finding which suggests that successful attainment of the white standard had high incentive value. This is plausible, given the extent to which the Negro college student's career prospects may depend upon his ability to meet white standards of competence. The apparent inconsistency between the self-reports of "caring" and actual performance in the two experimental conditions suggests that the white achievement standard might have been perceived not only as a more attractive goal, but also as a more difficult one. That is, in the white norm condition the depressant motivational effect of having a low expectancy of success could have outweighed the facilitating influence of the high incentive value of success. It follows that Negro men who are more confident of their ability to compete intellectually against white peers than were subjects in the Katz, Epps, and Axelson experiment may perform better with white norm instructions than with Negro norm instructions. Since the earlier study was done at a relatively nonselective Negro college in the Deep South, it was decided to carry out replications at two predominantly Negro colleges having higher admission standards. One institution was in the Deep South and the other was in the North.

Another topic of interest in connection with Negro test behavior is the influence on performance of two personality variables, hope of success and fear of failure, as inferred from n Achievement test scores alone, or in combination with test anxiety scores. Atkinson (1958), using arithmetic problems, and Feather (1965b), using anagrams, noted a tendency (significant in

---

[1] This research was supported by a grant (MH 08051-01) from the National Institute of Mental Health.

[2] Requests for reprints should be sent to Irwin Katz, City University of New York, Graduate Center, 33 West 42 Street, New York, New York 10036.

Feather's study) for highly success-motivated (high hope-of-success–fear-of-failure) white subjects to show optimal performance when the objective probability of success was moderately low, and for subjects with a relatively weak motive to succeed (low hope-of-success–fear-of-failure) to respond more favorably when the objective probability of success was high. A factor underlying this difference seems to be the higher self-confidence of persons with a strong disposition to succeed. Thus Brody (1963) and Feather (1965a) reported that high hope-of-success–fear-of-failure white subjects approached new achievement situations with higher expectancies of success than did persons with low hope-of-success–fear-of-failure scores, particularly when the attainment standard was moderately difficult. These various findings on white subjects have implications for the present study. Assuming that students from the two selective Negro colleges would perceive a white norm as moderately more difficult to attain than a Negro norm, any energizing effect of white norm instructions should be more apparent among subjects with relatively strong motive to seek success than among those with a relatively weak motive

EXPERIMENT I

The first experiment included, in addition to the two norm conditions, a neutral, or nonevaluative, condition. It was assumed that neutral instructions would be less motivating than aptitude test instructions. Hence, taking all three conditions into account it was predicted that the rank ordering of groups with respect to amount of pre–post gain in arithmetic scores would be as follows: white norm, Negro norm, neutral. It was also predicted that these differences between conditions would be larger for high-motive subgroups than for low-motive subgroups.

Method

Subjects and Instruments

The study was conducted at the same state-supported Negro college in the Deep South that had been used by Katz, Epps, and Axelson. But due to a change in the admission standards the subjects were of higher average academic ability. Whereas formerly all Negro high school graduates in the state were eligible to enroll, a new criterion excluded applicants falling roughly in the lower half of the state-wide distribution of Negro scores on a twelfth-grade aptitude test. Eighty-six male freshmen were used as subjects. The experimental task was arithmetic, chosen in preference to the code task of the earlier study because pilot research indicated it was perceived by students as a more credible test of intelligence. It consisted of 25 addition problems, each having three three-digit numbers arranged in vertical order for summing. The achievement motive was measured by means of a multiple-choice version of the French (1958) Test of Insight that was developed by Willerman, Lewit, and Tellegen (1960). Their achievement motive score is the number of hope-of-success responses minus the number of fear-of-failure responses. For the present study, 10 items unrelated to hope of success or fear of failure were eliminated from their questionnaire, leaving 11 hope of success and fear of failure items, and 4 filler items.

Experimental Procedure

One Negro male faculty member, previously unknown to the subjects, administered both the achievement motive questionnaire and the arithmetic task at two sessions held in freshman ROTC classes. At the first session all subjects received the same printed instructions. The achievement motive questionnaire, administered first, was titled "Viewing Social Behavior" and was described on the face sheet as part of a research project on "how people react to the behavior of others." Purportedly, replies had no academic significance and were to be treated as private. Next, the arithmetic task was administered, also with instructions that described it as a nonevaluative research instrument. Three minutes were allowed for this task.

The second session was held in the same classrooms a week later. Unknown to the subjects, about one-third of every class tested received a particular type of printed instructions. A no test group was given essentially the same instructions as they had received previously. But for two aptitude test groups the motive questionnaire was now titled "Social Intelligence Test," and described as "a test of social insight, which is an important component of general intelligence." For the two aptitude test groups the arithmetic task was now titled "Scholastic Aptitude Examination: Numerical Ability," and described as a measure of "quantitative intelligence–the ability to manipulate numbers quickly and accurately." On both tasks, the instructions also stated: "You recently were given a practice form of this test . . . to acquaint you with the procedure. Your score today . . . is the one that counts. It will be used to evaluate your intellectual ability by comparing you with other students."

The remainder of the instructions in the two aptitude test conditions contained the racial comparison variation. Subjects in the white norm condition read that they would be compared with all freshmen in the state university system. The other state university campuses were named. Since these were large institutions with virtually all-white enrollments, it was assumed that subjects would be aware that they were to be evaluated against essentially white standards of achievement. Subjects in the Negro norm condition read that they would be compared only with other freshmen at their own college. Again, working time on the arithmetic problems was 3 minutes. Then subjects filled out a questionnaire dealing with their reactions to the experiment, and were informed of the aptitude test deception.

Results

Adequacy of Procedures

Responses to an item in the terminal questionnaire indicate that the differential instructions tended to register and be remembered. Asked what was supposed to be done with their scores, over 85% of subjects in each condition correctly identified one of three alternatives: used for research only, compared with scores of own-college peers, compared with scores of all freshmen in the state university system.

Next, the reliability of the total achievement motive scores (hope-of-success–fear-of-failure) was assessed. The correlation of first and second scores for 30 subjects in the no test condition was .76, indicating adequate stability. Correlations for groups of similar size in the white norm and Negro norm conditions were .73 and .63, respectively.

Arithmetic Performance

Number of correct solutions. Two 3 × 2 analyses of variance were done on the number of arithmetic problems answered correctly, the independent variables being the three instructional conditions and a median split of first-session scores on the achievement motive questionnaire. First, an analysis of variance was performed on first-session arithmetic scores, as a check on the initial comparability of performance in the various treatment groups. There were no effects that approached significance.

Next, first-session to second-session gains in number of problems answered correctly were analyzed. Only the main effect for conditions was significant ($F = 3.40$, $df = 2/85$, $p < .05$). Table 1 reveals the following rank order of gain scores in the three instructional conditions: white norm, no test, Negro norm. When conditions are compared by means of the Newman-Keuls test (Winer, 1962) only the difference between the white norm and Negro norm means is significant ($p < .05$). There was also a trend toward an interaction effect of motive strength and instructions ($F = 2.62$, $df = 2/85$, $p < .10$). Comparing individual cells by the Newman-Keuls test indicates that in the high hope-of-success–fear-of-failure subsample the white norm occasioned significantly higher gains than either the Negro norm or the no test condition (for both comparisons, $p < .05$). On the other hand, there were no significant differences between any two low-motive groups.

TABLE 1

MEAN ARITHMETIC GAIN SCORES IN THREE CONDITIONS FOR SUBJECTS ABOVE AND BELOW MEDIAN OF FIRST-SESSION ACHIEVEMENT MOTIVE SCORES

Motive group	Condition					
	No test		White norm		Negro norm	
	M	*n*	*M*	*n*	*M*	*n*
High	1.45	13	3.93	14	1.54	13
Low	2.05	17	1.73	15	.01	14
Combined	1.83	30	2.78	29	.78	27

Thus the results provide limited support for the predictions. Subjects did not perform as expected in the neutral condition. However, comparing only the white norm and Negro norm conditions, both the total groups and the achievement motive subgroups tended to behave in accordance with the predictions.

Additional 3 × 2 analyses of variance were done on arithmetic scores, using conditions and second-session motive scores as independent variables. Again, first-session arithmetic scores in all cells were similar, justifying an analysis of variance of pre–post gain scores. This yielded, in addition to the conditions main effect already presented, a main effect of motive strength ($F = 4.50$, $p < .05$). Table 2 shows that high-

TABLE 2

MEAN ARITHMETIC GAIN SCORES IN THREE CONDITIONS FOR SUBJECTS ABOVE AND BELOW MEDIAN OF SECOND-SESSION ACHIEVEMENT MOTIVE SCORES

Motive group	Condition					
	No test		White norm		Negro norm	
	M	*n*	*M*	*n*	*M*	*n*
High	2.30	10	4.00	15	2.12	10
Low	1.60	20	1.50	14	.00	17

motive subjects performed better than low-motive subjects in all conditions.

Error scores. The four 3 × 2 analyses of variance described above were also done on arithmetic error scores (i.e., number of problems answered erroneously). Only one analysis yielded a significant result: using conditions and second-session motive scores as the independent variables, high-motive subjects showed more improvement in errors than low-motive subjects. (For the main effect of motive strength, $F = 4.10$, $p < .05$.)

Analysis of arousal effects. To evaluate the effect of experimental conditions on responses to the achievement motive questionnaire, one-way analyses of variance were done on pre–post changes in three types of score: hope of success minus fear of failure, hope of success alone and fear of failure alone. There were no significant results.

Additional Findings

Reported nervousness. Subjects were asked in the terminal questionnaire how nervous they had felt when working on the arithmetic problems. They answered on a 22-point scale ranging from extremely calm at the low end to extremely nervous at the high end. A 3 × 2 analysis of variance (Conditions × First-Session Motive Scores) revealed a significant main effect of conditions ($F = 3.38$, $p < .05$). The means were as follows: no test, 10.52; white norm, 8.80; and Negro norm, 12.34. Only the difference between white and Negro norms was significant by the Neuman-Keuls test ($p < .01$). When correlations were done within conditions between reported nervousness on one hand and either strength of achievement motive or measures of arithmetic performance on the other hand, no significant relationships were found.

Discussion

The position of the neutral condition subjects midway between the two norm samples on measures of arithmetic improvement and nervousness suggests that they were not lacking in motivation. As to the nature of the motivation, it can be noted that the neutral instructions contained an appeal for cooperation ("It will help the research if you work as fast and as accurately as you can") which may have aroused an expectancy of social approval for complying. This interpretation is consistent with a previously reported finding (Atkinson & Raphelson, 1956; French, 1955) of a positive relationship between n Affiliation scores and performance under a neutral condition similar to the one employed in this study. It is interesting to note that Katz, Epps, and Axelson's (1964) neutral condition subjects, who were not requested to work "fast and accurately," did not perform as well as subjects in either norm condition.

Looking only at the two aptitude test conditions in the present study, the data support the prediction of superior arithmetic performance by the white norm group. However, the finding of higher levels of reported nervousness among Negro norm subjects is somewhat anomalous. On the one hand, the nervousness scores appear to contradict the original assumption that Negro subjects would perceive a white intellectual standard as both more important and riskier than a Negro standard. But on the other hand, the lack of correlation between reported nervousness and either arithmetic scores or hope-of-success–fear-of-failure scores suggests that this variable was not related to motivation. If one makes the assumption that reported nervousness was in fact indicative of anxiety about failure, even though it did not correlate with other measures, it then becomes necessary to account for the high nervousness scores of the Negro norm subjects. Possibly the Negro student's awareness of racial inequalities in educational and other life opportunities enhance the anticipated gain in self-esteem attendant upon success in cross-racial competition, while making failure relatively easy to explain away. Hence, in the biracial achievement situation, as compared with the uniracial situation, the positive incentive value of success would have been greater, even though there was less anxiety about failure.[3]

Regarding the achievement motive, its effect on performance varied, depending on which set of hope-of-success–fear-of-failure scores were used. First-session motive scores showed a marginal tendency to interact with conditions, whereas second-session motive scores influenced performance independently of the conditions effect. However, the latter finding must be interpreted cautiously. Even though the aptitude test instructions did not significantly affect second-session motive scores, it should be recognized that the latter were not obtained under a single standard condition.

Comparison of this study with the previous one by Katz, Epps, and Axelson (1964) is difficult because the two experiments differed not only with respect to the academic ability of the subjects, but also with respect to the experimental task. To help clarify this interpretive ambiguity, as well as others mentioned above, another experiment was done in which the same task used by Katz, Epps, and Axelson was employed, and new efforts were made to assess (*a*) hope of success and fear of failure, and (*b*) subjects' perceptions of the attractiveness and difficulty of competing intellectually with own-race and other-race peers.

EXPERIMENT II

The main innovations in Experiment II were as follows: the neutral condition was dropped; in addition to the achievement motive, test anxiety was used as an independent variable; performance was measured at two tasks, scrambled words and digit-symbol substitution. It was predicted that the white norm condition would be more favorable for performance than the Negro norm condition. In view of the ambiguous findings in Experiment I on the effect of motive strength, no predictions were made regarding the two personality variables.

Method

The subjects of the experiment were 219 Negro male freshmen at a state-supported, predominantly Negro college in the North. The academic ability of the student body was similar to that of students at predominantly white campuses in the same state university system, as shown by published American College Testing Program scores.

Subjects were tested in groups of about 50 in two sessions that were conducted during ROTC class time. At the first session a Negro male member of the Psychology Department introduced himself and asked the men for their cooperation on a research project supposedly dealing with students' reactions to various tasks. Then the multiple-choice achievement motive scale and Sarason and Ganzer's (1962) test anxiety scale were administered. Next, each of two cognitive tasks was presented for 3 minutes. The first one consisted of scrambled five-letter words. Subjects were instructed to solve as many items as possible by writing in the correct words, until told to stop. The second task was the more difficult of two digit-symbol substitution tasks that had been used by Katz, Epps, and Axelson (1964).

About 1 week later, subjects were tested again by the same experimenter. The personality scales were not administered at this session. Equivalent forms of the two cognitive tasks were presented in the same order as before. Now each task had a face sheet with instructions that described it as a measure of scholastic aptitude. The scrambled words were described as "a test of ability to think quickly and accurately and to organize elements into meaningful wholes" and the code task was called "a test of ability to concentrate and to make quick and accurate mental responses." The instructions for accomplishing the race of norms variation were similar to those used in Experiment I.

After completing both tasks subjects filled out a questionnaire dealing with their reactions to the experiment. Finally, they were informed of the aptitude test deception, and the general purpose of the experiment.

Results

Scrambled Words Task

To find out whether the norm instructions affected performance on the scrambled

---

[3] This line of reasoning applies only to situations in which there is no actual contact with white peers or white adult evaluators. When such contacts are involved, the outcome also depends upon the Negro subject's expectations regarding white reactions to his achievement efforts (cf. Katz, 1964).

TABLE 3

Comparison of Performance of White Norm and Negro Norm Groups

Type of score	Norm group					t	p
	Negro[a]		White[b]		Diff. W − N		
	M	SD	M	SD			
Scrambled Words							
Preinstructions	17.91	5.59	18.29	6.21	.38	1.00	—
Pre–post change	14.66	7.32	17.66	7.58	3.00	2.97	.01
Code							
Preinstructions	60.85	16.70	59.62	14.49	−1.23	1.00	—
Pre–post change	19.63	9.00	22.96	8.44	3.33	2.38	.02

[a] $n = 107$.
[b] $n = 112$.

words, t tests were done on the number of items answered correctly. A comparison of first-session scores yielded a t value below 1.00, indicating that the two norm groups were not initially dissimilar in performance. Next, a t test was done on pre–post gain scores. As predicted, white norm subjects made significantly larger gains than Negro norm subjects ($t = 2.97$, $p < .01$). These comparisons are summarized in Table 3.

To evaluate personality effects, each motive variable (consisting of extreme tercile groups) was combined with the norm variable in a separate 2 × 2 analysis of variance. First-session scores on scrambled words showed no effects that approached significance. With regard to pre–post change scores, there was a main effect of test anxiety: low-anxious subjects showed larger performance gains than high-anxious subjects ($F = 4.20$, $df = 1/132$, $p < .05$). Also, there was a trend toward a main effect of the achievement motive, with the high-motive group making larger gains ($F = 2.95$, $df = 1/116$, $p < .10$).

Digit-Symbol Task

As can be seen in Table 3, when the two norm groups were compared on first-session code scores, their initial performance levels were similar, while comparison of pre–post change scores showed a significant advantage for white norm subjects, as predicted ($t = 2.38$, $p < .02$). When analysis of variance was used, no effects involving personality variables were obtained.

Additional Findings

In the terminal questionnaire subjects were asked whether they thought it would be more valuable to know how they compared in academic ability with all college students in the state, or with students at their own college. Over 80% indicated a preference for the statewide (i.e., predominately white) comparison referent. There were no significant differences associated with norm instructions or personality factors.

That white achievement standards were seen as somewhat more difficult to attain than Negro standards is suggested by replies to the question, "How well do you think you did on these tests, compared to other students who have taken them?" On the scrambled words task, subjects expecting to be compared with white students placed themselves in a lower mean percentile group (61.7) than did the Negro norm subjects (65.9). The difference was marginally significant ($t = 1.91$, $p < .07$). Comparisons of self and others on the code task revealed a smaller, nonsignificant effect of the norm variable in the same direction.

That the white norm for scrambled words, but not for code, tended to be seen as more difficult than the Negro norm is perhaps related to a difference in the meaning of the two tasks. Eighty-six percent of all subjects regarded scrambled words as "a better test of academic aptitude." This difference in perception may also account, in part, for the fact that the achievement motive and test anxiety variables influenced scores on scrambled words, but not on code. There was no effect of norm conditions on self-reports of nervousness in the experimental situation.

DISCUSSION

Experiments I and II show, as hypothesized, that when relatively simple cognitive tasks are presented by a Negro examiner, both northern and southern Negro male college students will respond favorably to an opportunity for intellectual comparison with white peers. By way of contrast, Katz, Epps, and Axelson had found that cross-racial competition was likely to have a detrimental effect on the cognitive performance of more severely disadvantaged Negroes (i.e., students at an academically nonselective Negro college in the Deep South).

The assumption underlying the hypothesis—that relatively capable Negro students would regard white achievement standards as more informative yet only moderately more difficult—appears to be supported by questionnaire data from Experiment II. However, subjects may have responded defensively to the questionnaire items. In this regard to anomalous data on reported nervousness are noteworthy. In Experiment I more nervousness was reported in the Negro norm condition, while in Experiment II there was no difference between norm groups in reported nervousness. It was hoped that the personality scales would help illuminate motivational process associated with the race-of-norm effect on performance. But in both experiments, whenever success-approach and failure-avoidant motives influenced task behavior their effects tended to be independent of the norm effect. Thus the psychological factors associated with cross-racial facilitation are still in need of clarification.

It would be worthwhile in future experiments to manipulate variables that seem to be related to the racial comparison effect, such as the objective probability of favorable comparison and the relevance of a given test score to the academic and vocational aspirations of subjects. At present one cannot predict with assurance the specific types of Negro students who may benefit from cross-racial competition, or the specific types of tasks and achievement settings that will reveal behavioral facilitation. A further limitation on the generality of the results is the fact that only Negro testers were used. Other studies of Negro college students (Katz & Greenbaum, 1963; Katz, Roberts, & Robinson, 1965) indicate that the race of tester variable can interact with other situational variables, such as degree of stress and probability of success, in affecting performance.

REFERENCES

ATKINSON, J. W. Towards experimental analysis of human motivation in terms of motives, expectancies, and incentives. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* New York: Van Nostrand, 1958.

ATKINSON, J. W., & RAPHELSON, A. C. Individual differences in motivation and behavior in particular situations. *Journal of Personality,* 1956, **24,** 349–363.

BRODY, N. n Achievement, test anxiety and subjective probability of success in risk taking behavior. *Journal of Abnormal and Social Psychology,* 1963, **66,** 413–418.

FEATHER, N. T. Performance at a difficult task in relation to initial expectancies of success, test anxiety, and need achievement. *Journal of Personality,* 1965, **33,** 200–217. (a)

FEATHER, N. T. The relationship of expectation of success to need achievement and test anxiety. *Journal of Personality and Social Psychology,* 1965, **1,** 118–126. (b)

FRENCH, E. G. Some characteristics of achievement motivation. *Journal of Experimental Psychology,* 1955, **50,** 232–236.

FRENCH, E. G. Development of a measure of complex motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* Princeton, N. J.: Van Nostrand, 1958.

KATZ, I. Review of evidence relating to effects of desegregation on the intellectual performance of Negroes. *American Psychologist,* 1964, **19,** 381–399.

KATZ, I., EPPS, E. G., & AXELSON, L. J. Effect upon Negro digit-symbol performance of anticipated comparison with whites and with other Negroes. *Journal of Abnormal and Social Psychology,* 1964, **69,** 77–83.

KATZ, I., & GREENBAUM, C. Effects of anxiety, threat, and racial environment on task performance of Negro college students. *Journal of Abnormal and Social Psychology,* 1963, **66,** 562–567.

KATZ, I., ROBERTS, S. O., & ROBINSON, J. M. Effects of task difficulty, race of administrator, and instructions on digit-symbol performance of Negroes. *Journal of Personality and Social Psychology,* 1965, **2,** 53–59.

SARASON, I. G., & GANZER, V. J. Anxiety, reinforcement, and experimental instructions in a free verbalization situation. *Journal of Abnormal and Social Psychology,* 1962, **65,** 300–307.

WILLERMAN, B., LEWIT, D., & TELLEGEN, A. Seeking and avoiding self-evaluation by working individually or in groups. In D. Willner (Ed.), *Decisions, values and groups.* Vol. 1. London: Pergamon Press, 1960.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received February 27, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 3, 209–214

THE MEANING OF PAIRED ASSOCIATES[1]†

JOEL R. LEVIN[2] AND JAMES M. HORVITZ
University of Wisconsin

A printed paired-associate list containing six different types of items was presented to 60 sixth graders. Four of the item types consisted of sentence-embedded paired associates on the study trials, followed by test trial cues varying in terms of their similarity in meaning to the study materials. The two remaining item types served as base-line measures of performance, that is, controls. It was found that (*a*) the greater the semantic similarity between the study and test trial contexts, the greater was the recall of paired-associate items, while at the same time (*b*) control item types presented in conjunctive phrases were better recalled than those presented in sentences.

In recent years the paired-associate paradigm has provided a framework for investigating the role that semantic and syntactic factors play in learning. For example, a great deal of empirical evidence suggests that the learning of a paired-associate list involves more than mere passive participation on the learner's part. Rather, subject reports obtained during and after learning, albeit introspective and less than perfectly reliable, indicate that proficient learners restructure or elaborate upon the to-be-associated items (by the addition of verbal or imaginal elements) in order to render the pairs more memorable (Martin, 1967; Paivio, 1969; Rohwer, 1967; Runquist & Farley, 1964).

Furthermore, it has been reported that experimenter-provided sentences or phrases, serving to relate paired-associate items to one another, facilitate the task (Davidson, 1964; Reese, 1965; Rohwer, 1966). That the kind of auxiliary verbal material supplied is important has also been confirmed. Rohwer (1966) found syntactic structure to be an effective facilitator of paired-associate learning only when the provided elaboration consisted of meaningful English words. Thus, sentences like "The running *cow* chases the bouncing *ball*" were facilitative, while Lewis Carrol–like pseudosentences ("The ludding *cow* drases the spraking *ball*") were not, relative to nonelaborated pairs (*cow—ball*).[3]

In a subsequent experiment, Rohwer and Levin (1968) discovered that the properties (*a*) meaningful, (*b*) English, and (*c*) words were necessary but not sufficient effectors of improved paired-associate performance. In that study, the way in which stimulus and response members were related to one another was crucial. When the relationship was semantically appropriate (e.g., *cats* jump *gates*) facilitation was produced; when semantically anomalous (e.g., *cats* jump *songs*), it was not.

The present study follows from these and other recent experiments (e.g., Davidson & Dollinger, 1969; Davidson, Schwenn, & Adams, 1970; Ehri & Rohwer, 1969) in which semantic characteristics of verbs have been manipulated in order to examine the relationship between sentence meaning and paired-associate recall.

---

[1] The research was conducted while the first author was affiliated with the Wisconsin Research and Development Center for Cognitive Learning. The authors acknowledge the cooperation of the staff and students of the Central Grade School in Stoughton, Wisconsin, as well as the helpful suggestions of Robert E. Davidson.

[2] Requests for reprints should be sent to Joel R. Levin, R & D Center for Cognitive Learning, The University of Wisconsin, 1404 Regent Street, Madison, Wisconsin 53706.

[3] This is interesting in light of Epstein's (1961) finding that nonmeaningful verbal material (e.g., "maff vlem ooth glox nerf") is more easily recalled when syntactic markers are inserted in generally agreed upon places (e.g., "The maff vlems oothly the glox nerfs"). Perhaps the distinction lies in the interpretability of the original learning materials: When they are essentially meaningless (cf. Epstein, 1961) the addition of verbal elements is facilitative; when they are basically meaningful (cf. Rohwer, 1966) the addition of verbal elements is beneficial only if they too are meaningful.

METHOD

Design and Materials

The learning materials, which were typed onto a memory drum tape, consisted of 12 pairs of familiar nouns embedded in five-word sentences or conjunctive phrases. The 12 items were randomly distributed throughout the list, appearing as one of six types, with 2 items apiece representing each type. The six item types were defined by the form in which they were presented to subjects on study and test trials. Four of these were included to test hypotheses concerned with semantic factors in paired-associate learning, while the remaining two served as controls.

Examples of the four experimental and two control item types may be found in Table 1. Beginning with E_1 and continuing through E_4, one will perceive a decrease in the semantic similarity between the sentences studied and the frames supplied as test cues. In E_1, for example, the verb initially studied is reinstated as a cue for recall on the test trial, while in E_2, a synonymous verb (or one which implies similar meaning) is provided. The extent to which the test verb evokes denotative and connotative responses different from those of the study might be manifested by performance differences on item types E_1 and E_2.

However, this difference should be smaller in magnitude than that anticipated when E_1 is compared with item types E_3 and E_4, where more than orthographical changes in the test-trial verb are involved. In E_3, the test verb signals an action or activity on the part of the stimulus noun different from what was communicated on the study trial. In E_4, on the other hand, the original meaning of the response noun has been completely altered by the verb switch: In the example in Table 1, what may have been encoded as "girl present at *gala event*" is now retrievable only from the cues "girl reaching or grabbing for something" (in this case, a *small spherical object*).

If semantic factors are involved in the learning of such paired-associate items, a decrement in performance would be predicted as a function of semantic dissimilarity (between study and test contexts). That is to say, in terms of recall the four item types would be rank-ordered as follows: $E_1 > E_2 > E_3 > E_4$, where at one extreme the study- and test-trial contexts are identical (E_1) while at the other they convey dissimilar, or even competing, sentence meaning (E_4). The characteristics of the predicted function will be empirically evaluated in the present study.

The two item-type controls, examples of which are also presented in Table 1, were incorporated to serve a two-fold purpose: (*a*) to provide sentence and nonsentence base lines against which to assess the contribution to learning of the hypothesized semantic factor; and (*b*) to yield an estimate of the previously discussed sentence facilitation effect for the present task, by comparing C_1 (sentence control) with C_2 (nonsentence control).

Each subject received one of six different combinations (arrangements) of item types and item content. This was counterbalanced across subjects in order that a partial test for possible item-content/type differences could be conducted.

TABLE 1

EXAMPLES OF STUDY- AND TEST-TRIAL MATERIALS

Item types	Study trial	Test trial
Experimental		
E_1	The *girl* grabs the *ball*	The *girl* grabs the ___
E_2	The *girl* takes the *ball*	The *girl* grabs the ___
E_3	The *girl* throws the *ball*	The *girl* grabs the ___
E_4	The *girl* attends the *ball*	The *girl* grabs the ___
Control		
C_1	The *girl* grabs the *ball*	The *girl* or the ___
C_2	The *girl* or the *ball*	The *girl* or the ___

Subjects

Sixty sixth-grade students from an elementary school serving a semirural community in the Midwest participated in the experiment. In addition, five other students were pilot tested in order that the experimenter might adapt to the experimental procedure. The names of all participants were randomly selected from the roster of sixth graders in the school.

Procedure

All subjects were individually tested. Upon entering the experimental room, the subject was seated in front of a Lafayette memory drum. The instructions were then read by the experimenter, followed by the presentation of three sample items on the memory drum, one at a time.

The sample items were given to familiarize the subject with the to-be-performed task, as well as to impress upon him the fact that for some of the items, study and test contexts would not be the same. This was done by means of two examples in which the connective was changed (verb to conjunction, and verb to preposition) from study to test trial, and one in which it was not (conjunction to conjunction). The order in which the sample items were presented was randomly predetermined for each subject.

Upon ascertaining that the subject understood the nature of the task, the experimenter presented the 12 study items to the subject one at a time, at a 4-second rate, while reading them aloud concurrently. Following the last item there was an 8-second intertrial interval, during which the experimenter readied the subject for the test trial by reminding him of the task. The test frames then appeared at a 4-second rate while read aloud by the experimenter, the subject's job being to supply (vocally) the missing response nouns. After the last test item was exposed, the subject was given another 8-second "reminder" interval, followed by another study trial–test trial cycle. On each of the two study and two test trials, the 12 paired-associate items appeared in a different (random) serial order, in order to remove position cues as a possible source of variance.

RESULTS

Learning was measured in terms of the number of correct responses given on each of the two test trials. A multivariate analysis of variance (for repeated measurements) was performed, in which the independent factor consisted of the particular arrangement of item types and content provided for the subject, while two other variables produced 12 within-subject measures: six item types in conjunction with two test trials.

The 12 dependent measures were initially transformed into 11 new variables in order that hypotheses regarding item-type differences, trials differences, and their interaction might be tested. In addition, the sum of the original 12 dependent measures provided a test for between-group (arrangements) differences in overall performance (Morrison, 1967).

All hypotheses in the present study were tested with the probability of a Type I error set equal to .05. The mean number of items correctly recalled on each trial, summed across the arrangements factor (which proved statistically nonsignificant), is presented for each item type in Table 2. A graphical representation of the item-type differences (across arrangements and trials) may be found in Figure 1.

TABLE 2

MEAN NUMBER OF ITEMS RECALLED ON EACH TEST TRIAL (OUT OF TWO) AS A FUNCTION OF ITEM TYPE

Item types	Trial 1	Trial 2	Sum	Difference
Experimental				
E_1	1.15	1.75	2.90	.60
E_2	.93	1.52	2.45	.59
E_3	.70	1.30	2.00	.60
E_4	.43	1.13	1.56	.70
Control				
C_1	.33	.97	1.30	.64
C_2	.70	1.40	2.10	.70
Trial totals	4.24	8.07		

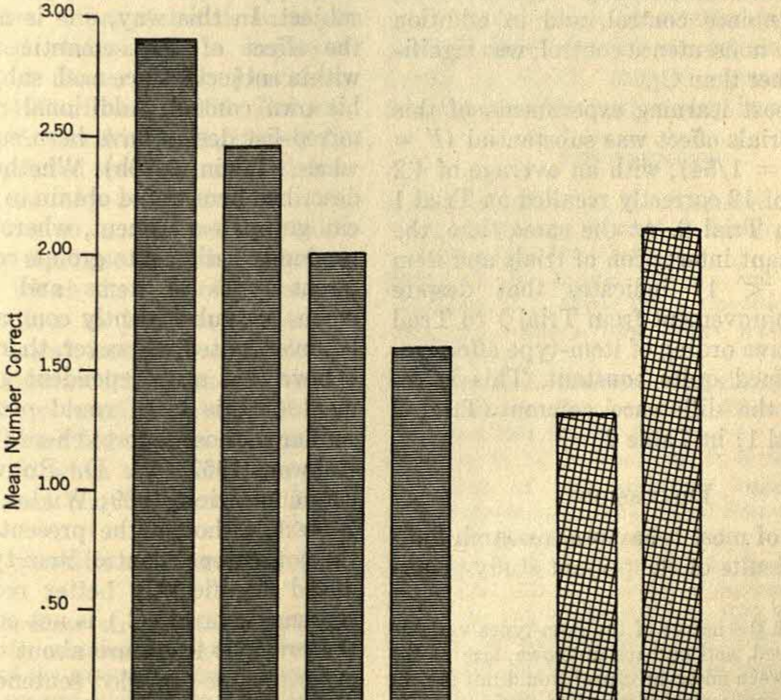


FIG. 1. Mean number of correct responses, by item type, over two test trials.

The multivariate test for item-type differences detected a statistically significant effect ($F = 21.10$, $df = 5/50$). Figure 1 reveals that the experimental item types conformed to the predicted pattern; that is, with decreasing semantic similarity between study and test contexts (going from E_1 to E_4) there is a corresponding decrement in response noun recall. Appropriate post hoc comparisons among the four experimental item-type means (Morrison, 1967) indicated that although differences between adjacent pairs were not statistically significant, E_1 was reliably different from E_3 and E_4, as was E_2 from E_4. More succinctly, as may be inferred from Figure 1 (and statistically corroborated by means of post hoc analysis), a negative linear trend accounts for virtually all of the variance in the difference among experimental item types.[4]

With respect to the control item types, it was found through post hoc analysis that (*a*) each of the experimental item types except E_4 (where a complete meaning change occurred betweeen study and test contexts) was statistically different from C_1, the sentence control, and in addition (*b*) C_2, the nonsentence control, was significantly higher than C_1.

As in most learning experiments of this type, the trials effect was substantial ($F = 327.98$, $df = 1/54$), with an average of 4.2 items out of 12 correctly recalled on Trial 1 and 8.1 on Trial 2. At the same time, the nonsignificant interaction of trials and item types ($F < 1$) indicates that despite marked improvement from Trial 1 to Trial 2, the relative ording of item-type effectiveness remained quite constant. This is reflected in the difference column (Trial 2 minus Trial 1) in Table 2.

DISCUSSION

Typical of most behavioral research findings, the results of the present study generate a host of questions still in need of resolution. A few of these will be intimated in the paragraphs which follow.

First, the data obtained from the four experimental item types attest to the importance of semantic factors in learning. What is more, the significant linear trend in recall as a function of the semantic similarity between study and test contexts emerged despite the fact that (*a*) there was an extremely restricted range of values for the number of correct items possible by item type (0–2 for each trial, 0–4 across trials); and (*b*) two trials were given, between which subjects were able to restructure inappropriate sets or better attune themselves to the "switch game" being played. Yet, the item-type profile observed on Trial 2 was virtually identical to that of Trial 1, with subjects exhibiting a consistent improvement of between .6 and .7 items (out of 2.00) on every item type (cf. Table 2).

In the present experiment a "mixed" list was incorporated to test the semantic hypothesis. That is, a single list containing each type of item was provided for every subject. In this way, one is able to assess the effect of the semantic manipulation within subjects, since each subject serves as his own control. Additional merits of the mixed-list design have been suggested elsewhere (Levin, 1970b). Whether the results described here would obtain in an independent groups experiment, where subjects are randomly assigned to groups containing different types of items and where group means are subsequently compared, needs to be investigated. However, there is reason to believe that an independent groups experiment of this kind would produce profiles similar to those reported here (Levin, 1970a; Rohwer, 1967; W. D. Rohwer, personal communication, 1969; Wilder, 1970).

Next, although the present finding that the nonsentence control item type (C_2) produced significantly better recall than the sentence control (C_1) is not consistent with the previous literature about the efficacy of experimenter-provided sentence elaboration (outlined earlier), it should be reexamined here for a few reasons. In the first place, the similarity of the contexts presented on study and test trials was greater for the nonsentence (conjunction-conjunction) than for the sentence (verb-conjunction) item types. In most of the experiments where the sentence effect has been obtained, the materials studied were generally sentences, conjunctive phrases, or noun-noun pairs followed by a test-trial cue of the stimulus noun by itself. In the present experiment, the switch from a sentence to a conjunctive phrase frame may have interfered with sentence/response noun retrieval, producing an effect equivalent to that observed when the sentence meaning was changed between study and test contexts. The fact that item types E_4 and C_1 were not statistically different lends support to this notion.

---

[4] Although the nature of the item-types variable (i.e., undefined, and perhaps unknown, size of the interval between adjacent types) would not permit an a priori application of trend analysis (see, for example, Winer, 1962), the present "linear trend" conclusion is simply based on the significant set of (linear) coefficients which best describes the data in a post hoc sense.

Of greater concern is that some recent studies employing printed materials have reported *not* detecting the now-famous sentence facilitation effect (Davidson, Schwenn, & Adams, 1970; Levin, 1970; Yuille & Pritchard, 1969), which was substantiated (or worse, detected *in the opposite direction*) in the present study. Since the phenomenon has been found to be most pronounced when either pictorial, aural, or subject-initiated elaboration is employed (e.g., Bobrow & Bower, 1969; Davidson, 1964; Paivio, 1969; Reese, 1965; Rohwer, Lynch, Levin, & Suzuki, 1967), it may be that provided printed sentence material does not facilitate, or may even interfere with, learning relative to nonsentence-provided material.[5] Why this is so might be suggested by a study in which the way of presenting verbal elaboration to learners—either in printed or in aural form, for example—is factorially manipulated. Such an experiment has currently been completed, where it was found that elaboration effects are indeed more pronounced when the verbalization is heard but not seen (Levin, Horvitz & Kaplan, 1971).

Finally, the question of imagery in children's learning may be at least alluded to with respect to the present results. If subjects did in fact engage in imaginal activity when attempting to encode the study pairs, the elicitation of the same image by means of an identical or similar context on the test trial might be assumed to facilitate response-noun retrieval relative to test contexts which were dissimilar or conflicting. Of course, this is what was observed among the experimental item types and interpreted in terms of *semantic* similarity.

Nevertheless, imaginal components of meaning need to be carefully isolated, if possible, and subjected to systematic investigation. For instance, subject reports regarding imagery usage might be related to the success with which item types of differing study–test semantic similarity are learned. Or, by inducing a child to employ imagery during the task by means of prelearning instructions followed by several examples (e.g., Bower, 1971; Spiker, 1960; Taylor & Black, 1969), one could relate the effect of this variable to the item-type profiles observed in this study (without imagery instructions). A comparison of these profiles might provide an estimate of the degree to which imagery occurs spontaneously in children when they learn.

REFERENCES

BOBROW, S., BOWER, G. H. Comprehension and recall of sentences. *Journal of Experimental Psychology*, 1969, **80**, 455–461.

BOWER, G. H. Mental imagery and associative learning. In L. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1971, in press.

DAVIDSON, R. E. Mediation and ability in paired-associate learning. *Journal of Educational Psychology*, 1964, **55**, 352–356.

DAVIDSON, R. E., & DOLLINGER, L. E. Syntactic facilitation of paired-associate learning: Deep structure variations. *Journal of Educational Psychology*, 1969, **60**, 434–438.

DAVIDSON, R. E., SCHWENN, E. A., & ADAMS, J. F. Semantic effects in transfer. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 212–217.

EHRI, L., & ROHWER, W. D., JR. Verb facilitation of paired-associate learning as a function of syntactic and semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 773–781.

EPSTEIN, W. The influence of syntactical structure on learning. *American Journal of Psychology*, 1961, **74**, 80–85.

---

[5] Although inferences of this kind frequently lend themselves to empirical validation (through an examination of the subject's incorrect overt responses), in the present study a determination of the nature of subject errors was not enlightening, since nearly three-quarters of the errors made under each item type were of the "omission" variety.

LEVIN, J. R. Factors related to the sentence facilitation of paired-associate learning: Some characteristics of verbs. *Journal of Educational Psychology*, 1970, **61**, 431–439. (a)

LEVIN, J. R. *Note on mixed lists in the analysis of repeated measurements*. (Technical Memo No. S-70-2) Madison: Wisconsin Research and Development Center, 1970. (b)

LEVIN, J. R., HORVITZ, J. M., & KAPLAN, S. A. Verbal facilitation of paired-associate learning: A limited generalization? *Journal of Educational Psychology*, 1971, in press.

MARTIN, C. J. Associative learning strategies employed by deaf, blind, retarded and normal children. *Educational Research Series*, 1967, **38**, 1–158.

MORRISON, D. F. *Multivariate statistical methods*. New York: McGraw-Hill, 1967.

PAIVIO, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–263.

REESE, H. W. Imagery in paired-associate learning in children. *Journal of Experimental Child Psychology*, 1965, **2**, 290–296.

ROHWER, W. D., JR. Constraint, syntax, and meaning in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 541–547.

ROHWER, W. D., JR. Social class differences in the role of linguistic structures in paired-associate learning: Elaboration and learning proficiency. (Final report on United States Office of Education Basic Research Project No. 5-0605. Contract No. OE 6-10-273) Washington, D.C.: United States Government Printing Office, 1967.

ROHWER, W. D., JR., & LEVIN, J. R. Action, meaning and stimulus selection in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 137–141.

ROHWER, W. D., JR., LYNCH, S., LEVIN, J. R., & SUZUKI, N. Pictorial and verbal factors in the efficient learning of paired associates. *Journal of Educational Psychology*, 1967, **58**, 278–284.

RUNQUIST, W. N., & FARLEY, F. H. The use of mediators in the learning of verbal paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 280–285.

SPIKER, C. C. Associative transfer in verbal paired-associate learning. *Child Development*, 1960, **31**, 73–87.

TAYLOR, A. M., & BLACK, H. B. Variables affecting imagery instruction in children. Paper presented at the annual meeting of the Midwest Psychological Association, Chicago, May 1969.

WILDER, L. Speech processes and verbal mediation in children. Research seminar, Wisconsin Research and Development Center for Cognitive Learning, Madison, February 1970.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

YUILLE, J. C., & PRITCHARD, S. Noun concreteness and verbal facilitation as factors in imaginal mediation and PA learning in children. *Journal of Experimental Child Psychology*, 1969, **7**, 459–466.

(Received March 4, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 3, 215–218

GROUP VARIABLES INFLUENCING THE TRANSFER OF CONCEPTUAL BEHAVIOR†

WILLIAM E. BEANE[1] AND ELMER A. LEMKE[2]
Illinois State University

Previous research has indicated that while groups can solve more problems than individuals, on subsequent individual transfer problems, the effect of group participation is lost. However, little attention has been given to group heterogeneity and ability level. In this study, 64 subjects were stratified by ability level and randomly assigned as homogeneous or heterogeneous pairs or quads to attain either three or seven training concepts. Following training, each subject participated individually on three transfer concepts. Analysis of transfer data indicates that training low-ability subjects as homogeneous quads facilitated individual transfer performance. These results and others are discussed.

Hudgins' (1960) study of the effects of group size on problem solving yielded the following conclusion—groups solve more problems in a training task than do individuals, although individuals *trained as individuals* solve more problems than individuals trained in groups in a subsequent transfer task. Although the conclusion has frequently been substantiated (Klausmeier, Wiersma, & Harris, 1963; Lemke, Randle, & Robertshaw, 1969), the conditions under which the conclusion holds have not been carefully delimited. Hudgins suggests that "... the problem solving superiority of small groups depends upon the efforts of the most able member of the group ... [p. 42]." Thus, intellectual heterogeneity of the group is posited as a factor in problem solving and is the focal variable in the present study.

Other factors known to affect transfer in conceptual tasks are also considered in the present study. Osler and Weiss (1962) have identified intelligence as an effective stratifying variable in concept learning (see also Denny, 1966; Freyberg, 1966; Osler & Fivel, 1961; Yudin, 1966). In addition, degree of initial training has been shown to have a significant effect on subsequent transfer (Grant & Berg, 1948; Grant & Cost, 1954).

The present study focused on the main and interactive effects of group structure (heterogeneous or homogeneous), general mental ability (high or low), group size (pairs or quads), and degree of training (three or seven concepts), on trials (1, 2, and 3) of the transfer concepts. Specifically, the following questions were of interest:

1. Will low-ability subjects trained in one of the two training conditions as homogeneous groups of pairs or quads attain transfer concepts as rapidly as low-ability subjects in similar training conditions in heterogeneous groups?
2. Will high-ability subjects trained in one of the two training conditions as homogeneous groups of pairs or quads attain transfer concepts as rapidly as high-ability subjects in similar training conditions in heterogeneous groups?

METHOD

Subjects

Subjects were 64 undergraduate students at Illinois State University. All subjects were in the age range 17–25. The mean Otis Gamma score of the group was 116.6, with a standard deviation of 8.47.

Materials

Stimulus materials forming the display consisted of 3 × 3 inch cards, each of which represented six attributes with two defining characteristics. The six two-valued attributes were as follows: border continuity (solid or broken), figure number (one or two), figure size (large or small), figure color (red or blue), figure position (high or low), and figure shape (circle or square).

The stimulus display consisted of an ordered arrangement of attributes, by rows and columns, which formed an 8 × 8 array of cards. The con-

---

[1] Now at Stout State University.

[2] Requests for reprints should be sent to Elmer A. Lemke, Department of Psychology, Illinois State University, Normal, Illinois 61761.

TABLE 1

MEANS FOR SIGNIFICANT SECOND-ORDER INTERACTION

Size × Structure × Ability	Homogeneous	Heterogeeous
High pairs	2.26	2.16
High quads	4.19	2.19
Low pairs	4.32	3.41
Low quads	2.18	3.52

cepts to be attained were three attribute conjunctive type concepts (e.g., large red squares).

Procedure

The Otis Quick-Scoring Mental Ability Test was administered to subjects during 1968–69, and the results were used to stratify subjects into high (GMA above 122) and low (GMA below 112) ability categories. Thirty-two high-ability subjects were randomly assigned to train as homogeneous or heterogeneous ability pairs or quads and to either three or seven training concepts. Similarly, 32 low-ability subjects were assigned as homogeneous or heterogeneous pairs or quads to one of the degree of training conditions. A heterogeneous group consisted of two subjects from each of the high- and low-ability categories; a homogeneous group consisted of four subjects from either the high- or low-ability categories. The 64 subjects made up 16 pairs and 8 quads. Following training, and to minimize waiting time for subjects, two members from each quad were randomly dismissed. Each of the two remaining quads and each member of the pairs participated individually in attaining three transfer concepts.

In the experimental setting, the experimenter randomly arranged the concepts each time naive subjects arrived. After instructing subjects on the structure of the stimulus display, the experimenter specified an exemplar of the concept as the focus card. As the subjects selected cards, for each choice the experimenter responded "yes," if the card was an exemplar of the concept and "no," if it was not. The subjects were told to cooperate in the task and specify the defining characteristics when they believed they had attained the concept. If the subjects attained a concept, the next was presented; if not, they continued for 10 minutes or until the concept was attained, whichever was shorter. After training, the experimenter recorded time-to-criterion scores for the individually attained transfer concepts. The results are based upon the analysis of these scores.

RESULTS

The analysis utilized a $2 \times 2 \times 2 \times 2 \times 3$ repeated measures summary ANOVA for group size, degree of training, group structure, ability level, and trials. In addition to the well-documented ability effect ($F = 5.00$, $df = 1/32$), the following four interactions were significant at the .05 level: (*a*) Size × Ability ($F = 4.46$, $df = 1/32$), (*b*) Size × Structure × Ability ($F = 4.45$, $df = 1/32$), (*c*) Size × Ability × Trials ($F = 4.40$, $df = 1/64$), and (*d*) Size × Training × Structure × Trials ($F = 4.14$, $df = 1/64$). The means for the significant second-order interactions are presented in Tables 1 and 2; the means for the first trial of the third-order interaction, the position in which the interesting interaction occurred, are presented in Table 3.

DISCUSSION

Previous investigation (Hudgins, 1960; Klausmeier et al., 1963; Lemke et al., 1969) has indicated that individuals trained as individuals perform better in a transfer setting than individuals trained in groups. However, the contemporary school practice of ability tracking and maintaining a small pupil-teacher ratio implies that students work well in smaller homogeneous groups. Hence, group size and homogeneity were the focus of the present investigation. The significant interactions confirm this focus and indicate some systematic patterns in research on classroom grouping.

Lemke and Hecht (1969) have suggested that subjects trained as low-homogeneous pairs each bring information which is shared and processed, thus facilitating individual performance on the transfer task. The present findings confirm this result (Table 1) and are consistent with results of classroom studies which indicate that it is the low-ability student who benefits from homogeneous grouping (Edminston & Benfer, 1949; Wilhelms & Westby-Gibson, 1961). Amaria, Biran, and Leith (1964)

TABLE 2

MEANS FOR SIGNIFICANT SECOND-ORDER INTERACTION

Size × Ability × Trials	Trial 1	Trial 2	Trial 3
High pairs	2.21	2.35	2.07
High quads	4.60	2.85	2.11
Low pairs	4.46	4.01	3.13
Low quads	2.03	3.17	3.37

find heterogeneous pairing a facilitating condition for the transfer performance of each pair member. This finding is also consistent with the present results (Table 1) and with early classroom studies which indicate that high-ability subjects benefit from being grouped heterogeneously (Billett, 1928; Hartill, 1936). While high-ability subjects appear to benefit from heterogeneous grouping (Table 1), training high-ability subjects as homogeneous quads inhibited individual transfer performance.

Basic transfer principles can be used to explain the above conclusions. Accept the principle that some strategy is required to systematically attain the transfer concepts. Since the results are reasonably conclusive (Hudgins, 1960; Klausmeier et al., 1963) that grouping subjects increases training efficiency, it follows that subjects have probably adopted a group strategy.

From the previously cited effects of intelligence we can infer that when low-ability subjects come to a training task, the probability is reasonably low that either will quickly develop a systematic individual strategy. Thus, they cooperatively develop a group strategy which can be modified in the individual performance of the transfer setting—proactive facilitation occurs. The more low-ability subjects in the group, the more information is processed, the greater the facilitation (Table 1, 2.18 < 4.32).

When four high-ability subjects attempt to train as homogeneous quads, the probability is higher that each subject will quickly develop an individual strategy. Group interaction forces the subject to adopt an alternate group strategy because high-ability subjects each develop and bring a different individual strategy to the setting. A great deal of interference occurs as subjects impose and/or develop a group strategy. When the high-ability subject again finds himself in the individual transfer setting he requires the old individual strategy or a variant thereof. As the high-ability subject attempts to develop an individual strategy, retroactive inhibition occurs (Table 1, 4.19 < 2.19).

When subjects train as heterogeneous pairs, low-ability subjects adopt the strategy of the high-ability member. In the individual transfer task, the performance of low-ability subjects trained as heterogeneous pairs is seen to be superior to that of low-ability subjects trained as homogeneous pairs—the result of adopting the strategy of the high-ability pair member. Proactive facilitation is seen to occur (Table 1, 3.41 < 4.32).

TABLE 3

MEANS FOR THE FIRST TRIAL

Size × Training × Structure	Training concepts	
	Three	Seven
Homogeneous pairs	4.27	2.65
Homogeneous quads	3.11	4.39
Heterogeneous pairs	3.43	2.84
Heterogeneous quads	4.88	.89

Low-ability subjects trained in heterogeneous quads do not perform as well in the individual transfer setting as low-ability subjects trained in homogeneous quads (Table 1, 3.52 < 2.18). The inference here is that the two high-ability subjects attempting to impose their individual strategy on the remaining quad members creates interference.

In conclusion, when ability level and group variables are controlled, the effects of group training on individual transfer performance are complex but orderly. While the ability effect obtained with college students corresponds closely with results obtained with children of elementary school age, the further interactive interpretations may not be generalizable to the younger population. Certainly the effects discussed have implications for small group training and study at the college level.

REFERENCES

AMARIA, R., BIRAN, L., & LEITH, G. Individual versus co-operative learning: I. Influence of intelligence and sex. *Educational Research,* 1964, **11,** 95–103.

BILLETT, R. A controlled experiment to determine the advantages of homogeneous grouping. *Educational Research Bulletin,* 1928, **7,** 133–140.

DENNY, J. Effect of anxiety and intelligence on concept formation. *Journal of Experimental Psychology,* 1966, **72,** 596–602.

EDMISTON, R., & BENFER, J. The relationship between group achievement and range of abilities within the groups. *Journal of Educational Research,* 1949, **42,** 547–548.

FREYBERG, P. Concept development in Piagetian terms in relation to school attainment. *Journal of Educational Psychology,* 1966, **57,** 164–168.

GRANT, D., & BERG, E. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology,* 1948, **38,** 404–411.

GRANT, D., & COST, J. Continuities and discontinuities in conceptual behavior in a card sorting problem. *Journal of Genetic Psychology,* 1954, **50,** 237–244.

HARTILL, R. *Homogeneous grouping as a policy in the elementary schools in New York City.* (Contributions to Education No. 690) New York: Columbia University, Teachers College, 1936.

HUDGINS, B. Effects of group experience on individual problem solving. *Journal of Educational Psychology,* 1960, **51,** 37–42.

KLAUSMEIER, H., WIERSMA, W., & HARRIS, C. Efficiency of initial learning and transfer by individuals, pairs, and quads. *Journal of Educatioinal Psychology,* 1963, **54,** 160–164.

LEMKE, E., & HECHT, J. Effect of degree of training, group size, and ability level on transfer of conceptual behavior. Unpublished manuscript, Illinois State University, 1969.

LEMKE, E., RANDLE, K., & ROBERTSHAW, S. Effects of degree of initial acquisition, group size, and general mental ability on concept learning and transfer. *Journal of Educational Psychology,* 1969, **60,** 75–78.

OSLER, S., & FIVEL, M. Concept attainment: I. The role of age and intelligence in concept attainment by induction. *Journal of Experimental Psychology,* 1961, **62,** 1–8.

OSLER, S., & WEISS, S. Studies in concept attainment: III. Effect of instructions at two levels of intelligence. *Journal of Experimental Psychology,* 1962, **63,** 528–533.

WILHELMS, F., & WESTBY-GIBSON, D. Grouping research offers leads. *Education Leadership,* 1961, **18,** 410–413, 476.

YUDIN, L. Formal thought in adolescence as a function of intelligence. *Child Development,* 1966, **37,** 697–708.

(Received March 8, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 3, 219–228

TEACHER APTITUDE AND OBSERVATIONAL LEARNING OF A TEACHING SKILL[1]†

MARY LOU KORAN[2]
University of Texas

RICHARD E. SNOW
Stanford University

AND FREDERICK J. McDONALD
New York University

In an experiment designed to investigate individual differences in acquiring a teaching skill from written and video-mediated modeling procedures, 120 subjects were given aptitude tests representing verbal and perceptual abilities, then randomly assigned to a video-modeling, written-modeling, or no modeling treatment. Criterion performances were the frequency, variety, and quality of the modeled teaching skill used on three separate microteaching sessions and on three written measures. Analysis of variance results showed that video modeling produced significantly higher performance frequencies than written modeling, with both modeling treatments significantly outperforming the control treatment. Regression analyses of Aptitude × Treatment interactions showed that Hidden Figures, Maze Tracing, Film Memory, and Sentence Reproduction scores interacted significantly with the modeling treatments.

The hypothesis that learning effectiveness is a function of the interaction of instructional treatments and individual-difference variables has not yet been widely tested. However, scattered research does indicate that subjects may learn more easily from one method than another, that this best method differs from subject to subject and that such differences may be related to learner characteristics (Cronbach & Snow, 1969).

One form of learning in which the role of learner characteristics has not been explicitly studied is that of observational learning. This literature (Bandura & Walters, 1963; McDonald & Allen, 1967) suggests the educational significance of studying learning in association with modeling procedures. Research in this area has consistently shown that complex behavior may be acquired or modified through observation with no direct external reinforcement. While it has been suggested that observer characteristics influence the extent to which observational learning occurs, observer characteristics have not been systematically studied. Such research is needed to evaluate factors influencing learning efficiency when modeling procedures are used for educational purposes, as in teacher training programs.

The purpose of this study was to examine the effects of verbal and perceptual aspects of teacher aptitude, in relation to two kinds of modeling presentations, on the acquisition of a teaching skill. The treatment variable employed was written modeling versus video-mediated modeling. While both treatments have been effective as training procedures (Bandura & Mischel, 1965; McDonald & Allen, 1967), differences between the two treatments have not been consistently significant. In view of task differences between the treatments, different abilities may be involved in each. Both treatments were expected to produce greater behavior change than a control treatment. It was anticipated that for subjects receiving video-mediated modeling, criterion scores should show stronger relation to perceptual abilities, while for subjects receiving written modeling, criterion scores should show stronger relation to verbal abilities. Such hypotheses imply nonparallel regression slopes, and consequently, that one treatment will not be superior throughout the aptitude distribution. These predictions were derived from an analysis of task and ability variables based on Melton's (1967) theoretical model for investigating individual differences in learning.

---

[1] This research was supported by a grant from the United States Office of Education, OEG-9-8-081073(010) and was an affiliated project of the Stanford Center for Research and Development in Teaching.

[2] Requests for reprints should be sent to Mary Lou Koran, Department of Educational Psychology, The University of Texas at Austin, Austin, Texas 78712.

Melton distinguished three components in associative learning: stimulus differentiation, representing the subject's coding response to the physical stimulus; association, representing direct or mediated linkage between the encoded stimulus and the required response; and response integration, representing the required output response or combination of responses. In terms of the stimulus differentiation component of Melton's model, subjects in the video-mediated modeling view a portrayal of the teaching skill at the actual pace of the lesson. This requires processing information from multiple channels simultaneously. Video modeling is rich in perceptual detail, with a complex mixture of relevant and irrelevant cues. These task characteristics would appear to require attention to task-relevant features despite perceptual distraction. Since video-modeling subjects must perceive and encode events as behavioral representation at a rate forced by video presentation, they can review relevant material in memory only. Thus the task should involve speed in perceptual evaluation and in exploring complicated spatial fields as well as considerable short-term memory facility. In written modeling, subjects read a written transcript of the sound track from video modeling. Thus, written-modeling subjects process information from a single channel and can establish their own pace for reading, selective scanning, and review. While subjects in written modeling must also distinguish significant features of the model, the stimulus situation is completely verbal. Thus the task should require ability to comprehend written verbal material.

Corresponding to the mediational component of the model, it is hypothesized that subjects in both modeling treatments must generate their own analog of the model's performance. While video-modeling subjects abstract relevant verbal representation from the behavioral presentation, written-modeling subjects must generate relevant behavioral representation from the verbal script. Accordingly, transformation between verbal and behavioral representation systems seems required.

Finally, in the response-integration component of the model, subjects in both treatments would be expected to retain representations of their respective models, generate new examples of the skill being modeled, and integrate these examples into a behavioral context. Verbal fluency would seem required as would the ability to remember both major ideas and details, from audiovisual presentation for video-modeling subjects and written presentations for written-modeling subjects. The ability variables chosen through this analysis are shown in Table 4. A more detailed presentation of the treatment-task analysis has been provided by Koran (1969).

METHOD

Subjects

A total of 121 subjects were drawn from among the intern teachers of the Stanford Teacher Education Program. The subjects were stratified by subject matter specialty, then randomly assigned to treatment conditions. All subject-matter specialties were represented among the teachers except foreign languages and physical education, which were believed inappropriate for the planned training.

Procedure

Ability measures were selected from the Educational Testing Service's Kit of Reference Tests for Cognitive Factors (French, Ekstrom, & Price, 1963) and from a series of film and audio tests (Seibert, Reid, & Snow, 1967; Seibert & Snow, 1965). Following test administration, subjects were randomly divided among three treatment groups: Video modeling consisted of videotape portrayal of the particular teaching skill to be learned; written modeling consisted of verbatim text from the sound track of the video model; no model represented a control treatment with the same procedural and practice steps but no observational treatment. Thus treatments were identical except for the mode of model presentation. All subjects received the same initial instructions and microteaching pretest, followed by two cycles of model presentation, rehearsal, and microteaching. Microteaching consisted of a laboratory arrangement in which subjects taught a prepared lesson to a group of four high school students. The subjects taught a different group of students in each lesson. This permitted them to revise the same basic subject matter while attending to the improvement of the teaching skill portrayed. Three written criterion measures completed the treatments. The procedure is summarized in Table 1.

Dependent Variables

The teaching skill portrayed in the modeling presentation was termed "analytic questioning." Such teacher questions are designed to engage students in the analysis of their textual reading material. The basic idea derives from the taxonomy by Bloom (1956) and his colleagues from which the general objective of analysis was selected as a basis for training teachers in questioning behavior.

Both classroom performance and written criterion measures of analytic questioning skill were used to assess treatment effects and ability–performance relationships. Classroom performance measures from each microteaching session included the frequency of analytic questions asked, categories of analytic questioning used, and quality of analytic questioning used as measured by similarity to behavior exhibited by the model (Koran, 1969). Typed transcripts of the microteaching sessions were independently rated for these variables by four judges, with interrater agreement ranging between .81 and .97 as determined by analysis of variance (Winer, 1962, pp. 124–132). Written criterion measures were (*a*) a recall test in which subjects were asked to generate a written list of analytic questions based on their previous lesson material; (*b*) an identification test, where subjects identified major categories of analytic questioning; and (*c*) a matching test, where subjects matched questions according to membership in a particular category of analytic questioning. The recall measure was rated in a manner identical to that of the classroom performance measures.

RESULTS

Because of space limitations, tests on simple main effects, separate comparison tests, and complete regression results have not been included here. Additional analyses are provided by Koran (1969).

Treatment Effects

A 3 (treatment groups) × 3 (teaching sessions) repeated measures analysis of variance was used to test instructional treatment main effects for classroom performance measures. Since conformity of the data to assumptions underlying the use of this model appeared questionable in some instances, separate one- and two-way analyses of variance were also computed for each session and each measure of the dependent variable (Koran, 1969). Since results were the same for both methods of analysis, the repeated measures analysis of variance was used in reporting these data. One-way analysis of variance was used to compare treatment groups on the written measures, since there were no repetitions of these. Due to presence of significant interactions, tests of simple main effects were computed, and the Newman-Keuls procedure was used to compare individual cell means following overall significant *F* ratios (Winer, 1962). Table 2 presents means and standard deviations of all measures of the dependent variables for each treatment group.

TABLE 1

TREATMENT PROCEDURES

Steps	Treatment group			Time
	Video modeling	Written modeling	No modeling	
Set induction	X	X	X	5 minutes
Teach 1 (Pretest)	X	X	X	10 minutes
Model 1	Video-mediated	Written	Extraneous material	10 minutes
Rehearsal 1	X	X	X	5 minutes
Teach 2	X	X	X	10 minutes
Model 2	Video-mediated	Written	Extraneous material	10 minutes
Rehearsal 2	X	X	X	5 minutes
Teach 3	X	X	X	10 minutes
Written criterion Test administration	X	X	X	15 minutes

Note.—Explanation of symbols: X indicates that all subjects received this step of the treatment in an identical manner. Written descriptions are provided for the two steps in which treatments varied among the three groups. Extraneous material was taken from Skinner (1968).

Table 3 presents analysis of variance results. These analyses strongly support the hypothesis that both modeling treatments would produce greater behavior change than the control treatment. On classroom performance measures, highly significant treatment effects were found for frequency, variety, and quality of analytic questioning. Comparisons between pairs of treatments showed that video modeling led to significantly higher frequency, variety, and quality of analytic questioning than did written modeling ($p < .01$) and that both modeling treatments led to significantly higher performance frequencies than the no modeling treatment ($p < .01$). This was true for Teaching Sessions 2 and 3. Between-group differences were not found for

TABLE 2

MEANS AND STANDARD DEVIATIONS OF DEPENDENT VARIABLES

Criterion measure	Teaching session	Video modeling[a] M	Video modeling[a] SD	Written modeling[b] M	Written modeling[b] SD	No modeling[c] M	No modeling[c] SD
Classroom performance measures							
Frequency of analytic questions	1	3.61	3.37	3.62	3.30	3.72	3.10
Frequency of analytic questions	2	13.40	6.41	8.16	5.59	3.31	2.50
Frequency of analytic questions	3	13.58	6.33	10.09	7.11	2.93	3.25
Categories of analytic questions	1	1.43	1.12	1.75	1.01	1.76	1.11
Categories of analytic questions	2	3.73	1.19	2.89	1.39	1.59	1.05
Categories of analytic questions	3	3.85	1.17	2.98	1.37	1.33	.97
Frequency of high quality analytic questions	1	1.88	2.02	1.75	1.89	2.18	2.27
Frequency of high quality analytic questions	2	9.90	5.20	5.17	3.84	1.92	1.70
Frequency of high quality analytic questions	3	10.15	5.52	6.72	4.42	1.68	2.24
Written measures							
Recall test							
Frequency of analytic questions		10.31	3.98	8.19	4.36	2.61	2.59
Categories of analytic questions		4.17	1.03	3.63	1.37	1.45	1.03
Frequency of high quality analytic questions		7.08	3.17	4.40	2.76	1.44	1.70
Matching test score		11.82	2.36	10.67	3.90	7.97	2.88
Identification test score		10.70	2.02	9.77	2.35	7.00	1.90

[a] $n = 41$.
[b] $n = 40$.
[c] $n = 39$.

Teaching Session 1, the pretest. Similarly, results of analyses of variance computed on scores for the written recall measure yielded significant treatment effects for frequency ($F = 46.19$, $df = 2/117$, $p < .01$), variety ($F = 63.55$, $df = 2/117$, $p < .01$), and quality ($F = 37.62$, $df = 2/117$, $p < .01$) of analytic questioning. Both video modeling and written modeling led to significantly higher performance frequencies than did the no modeling treatment ($p < .01$). Again, the two modeling procedures proved to be differentially effective, with video modeling subjects producing significantly higher frequency, variety, and quality of analytic questioning than did subjects in the written modeling treatment ($p < .01$).

Highly significant treatment effects were also found for scores on both the identification test ($F = 20.76$, $df = 2/117$, $p < .01$) and the matching test ($F = 32.94$, $df = 2/117$, $p < .01$). For both tests, subjects in the written- and video-modeling treatments performed significantly better than did subjects in the no modeling treatment ($p < .01$). Video modeling was superior to written modeling on the matching test ($p < .05$), but these two treatments did not differ on the identification test.

The effectiveness of the modeling treat-

TABLE 3

ANALYSIS OF VARIANCE FOR PERFORMANCE CRITERION MEASURES

Source	df	Frequency of analytic questions MS	Frequency of analytic questions F	Categories of analytic questions MS	Categories of analytic questions F	Frequency of high quality analytic questions MS	Frequency of high quality analytic questions F
Between subjects	119						
Treatment groups	2	1422.66	32.10*	64.79	24.75*	874.83	38.67*
Subjects within groups	117	44.32		2.36		22.62	
Within subjects	240						
Teaching sessions	2	992.67	75.31*	47.14	54.17*	614.88	79.74*
Treatment × Sessions	4	387.64	29.41*	24.06	27.65*	251.48	32.61*
Sessions × Subjects within groups	234	13.18		.87		7.72	

* $p = .01$.

ment was also shown by within-group analyses of change in analytic questioning behavior across sessions. Both modeling conditions produced significant increases in the frequency, variety, and quality of analytic questioning ($p < .01$), while the no modeling group did not display such increments (Table 3).

Aptitude × Treatment Interactions

For the combined groups, a correlation matrix of all aptitude and criterion variables is shown in Table 4. Variables were not combined by factor analysis despite some instances of high intercorrelation. Aptitude × Treatment interactions were evaluated by comparing regression slopes obtained for each aptitude-criterion pair in different treatments using F tests for heterogeneity of regression. Part scores for aptitude tests were analyzed separately because earlier work with some of these tests had indicated that the parts functioned differently as predictors of learning criteria. A total of 73 regression tests resulted in 13 significant Aptitude × Treatment interactions, using the .05 level of significance. The significant results are summarized in Table 5. Although several such interactions were obtained, the direction of the interactions did not consistently correspond to predictions and were at times opposite to the direction hypothesized.

Analyses of Aptitude × Treatment interactions indicated that scores on Film Memory interacted significantly with presentation conditions for the quality of analytic questioning on measures of classroom performance, and for both frequency and quality of analytic questioning on the written recall measure. In each case, Film Memory scores were positively related to performance in video modeling, while unrelated to performance in written modeling. In addition, scores on Sentence Reproduction, Part 2, interacted significantly with presentation conditions for the categories of analytic questioning used on the written recall meas-

TABLE 4

Correlations of Ability Measures with Performance Measures—Total Sample

Ability measures	Teaching Session 1			Teaching Session 2			Teaching Session 3			Recall test			Identification test	Matching test
	Analytic	Categories	High quality	Analytic	Cate gories	High quality	Analytic	Categories	High quality	Analytic	Categories	High quality		
Hidden Figures														
Part 1	17*	17*	16*	06	03	04	10	00	07	04	08	03	03	03
Part 2	−04	−03	−04	04	−05	01	03	−05	02	05	07	07	11	02
Total	08	08	07	06	−01	02	07	−03	05	05	08	05	08	06
Identical Pictures														
Part 1	16*	28**	13	01	04	04	05	03	02	06	12	01	06	15*
Part 2	20*	27**	18*	07	04	09	09	06	08	13	12	09	14	12
Total	20*	29**	17*	05	03	08	07	05	06	12	12	17*	11	14
Maze Tracing														
Part 1	08	17*	10	−05	−02	−06	01	05	−04	−04	−06	−12	−01	−07
Part 2	12	20*	11	−06	−09	−06	01	03	00	−06	−12	−13	00	04
Total	10	20*	11	−06	−06	−07	01	04	−02	−06	−10	−14	00	−02
Word Arrangement														
Part 1	10	14	08	−05	−08	−08	04	00	01	02	04	01	11	05
Part 2	17*	13	13	10	00	11	13	12	14	03	03	00	20*	09
Total	15*	15*	08	03	−04	02	09	07	08	03	04	00	17*	07
Advanced Vocabulary														
Part 1	23**	26**	13	11	03	07	03	−02	—04	10	07	−02	−01	01
Part 2	21**	23**	08	18*	11	12	08	02	00	11	12	01	11	12
Total	24**	26**	12	14	07	09	06	−01	−03	10	10	−02	06	06
Verbal Graduate Record Examination	25**	20*	09	19*	13	14	12	05	06	17*	21**	05	15*	23**
Film Memory	−04	−01	−07	05	07	01	17*	20*	17*	13	09	19*	09	12
Memory for Ideas														
Part 1	18*	20*	14	14	12	11	10	07	09	13	11	08	04	13
Part 2	20*	14	13	18*	15*	16*	22**	22**	20*	22**	18*	10	12	16*
Total	22**	22**	16*	17*	14	15*	16*	15*	15*	16*	15*	09	12	19*
Sentence Reproduction														
Part 1	07	06	05	13	00	06	02	—04	−03	22**	25**	14	08	15*
Part 2	13	03	01	20*	10	20*	12	04	12	15*	10	06	09	17*
Total	11	04	03	19*	06	16*	08	01	05	22**	21**	12	10	17*

Note.—Decimals omitted. $N = 120$.
* $p < .05$.
** $p < .01$.

ure. Scores on Sentence Reproduction were positively related to video-modeling performance while unrelated to written-modeling performance. These results are consistent with initial predictions.

In contrast to these findings, however, scores on Hidden Figures, Part 1, produced significant disordinal interactions for the frequency, variety, and quality of analytic questioning on classroom performance measures, with the magnitude of the interactions increasing across teaching sessions. Scores on Hidden Figures were positively related to written-modeling performance, but negatively related to video-modeling performance.

Similarly, for performance on the identification and matching tests, Film Memory and Maze Tracing, Part 2, scores interacted significantly with modeling presentations. Although Film Memory scores interacted significantly with scores on the identification test, the direction of differences was opposite to those previously reported for Film Memory scores. Here, the slope obtained for written modeling was positive, while a negative relation was obtained between aptitude and outcome in video modeling. Maze Tracing scores were positively related to written-modeling performance while unrelated to video-modeling performance.

All reported interactions were disordinal. Plots of the data showed in each case that regression lines for video modeling and written modeling intersected somewhere within the range of the aptitude variable in question. Such findings imply in general that assigning subjects on opposite sides of the intersection to alternative treatments is practically worthwhile. Thus, for example, subjects with relatively high scores on Hidden Figures, Part 1, would be expected to benefit more from the written-modeling treatment than from video modeling, while those with relatively low scores on Hidden Figures, Part 1, would probably benefit more from video modeling than from written modeling. In the present case, the strong main effect showed the video-modeling treatment superior to written modeling through most of each single aptitude range. There were actually few individuals for whom written modeling was the superior treatment. This lack of clear treatment differences at both ends of the aptitude con-

TABLE 5

SUMMARY OF SIMPLE REGRESSION ANALYSES SIGNIFICANT DISORDINAL INTERACTIONS

Criterion	Teaching session	Predictor			Treatment groups						F
		Aptitude test	M	SD	Video modeling		Written modeling		No modeling		
					a	b	a	b	a	b	
Analytic questions	2	Hidden Figures, Part 1	6.31	3.77	14.90	−.24	4.19	.61**	3.34	.02	4.02*
Analytic questions	3	Hidden Figures, Part 1	6.31	3.77	16.08	−.39*	3.87	.96**	2.71	.05	8.25**
Categories of analytic questions	2	Hidden Figures, Part 1	6.31	3.77	4.11	−.06	2.14	.12*	1.76	−.04	3.28*
Categories of analytic questions	3	Hidden Figures, Part 1	6.31	3.77	4.76	−.14**	2.08	.14**	1.30	.01	8.88**
High quality analytic questions	2	Hidden Figures, Part 1	6.31	3.77	11.54	−.25*	2.57	.40*	1.63	.04	4.20*
High quality analytic questions	3	Hidden Figures, Part 1	6.31	3.77	11.89	−.27*	3.01	.57**	1.88	−.03	5.70**
High quality analytic questions	3	Film Memory	20.38	3.98	−2.38	.62**	7.16	−.02	−.85	.13	4.42*
Recall test											
Analytic questions		Film Memory	20.38	3.98	1.83	.42*	8.18	.00	4.05	−.07	3.11*
Categories of analytic questions		Sentence Reproduction, Part 2	140.23	25.67	2.39	.01**	4.98	−.01	.09	.01	3.51*
High quality analytic questions		Film Memory	20.38	3.98	−3.50	.41**	3.01	.05	2.12	−.04	3.47*
Identification test		Film Memory	20.38	3.98	13.09	−.11	6.32	.17*	3.65	.17*	4.01*
Identification test		Maze Tracing, Part 2	14.06	4.03	10.74	.00	5.91	.27**	8.58	−.10	4.94**
Matching test		Maze Tracing, Part 2	14.06	4.03	12.95	−.08	3.77	.47**	9.09	−.07	7.44**

Note.—a and b are of the form $\hat{Y} = a + bx$.
* $p = .05$.
** $p = .01$.

tinua is not regarded as critical at this stage of Aptitude × Treatment interaction research, for two reasons. First, these initial Aptitude × Treatment interactions serve only as working hypotheses to be improved upon by further refinement of both aptitude and treatment variables. Second, more sophisticated statistical analyses, including a consideration of cost factors, may capitalize on aspects of such data not apparent in the present analyses. While a complete examination of these further approaches is beyond the scope of this study, two examples may be included to suggest possible directions for future work.

Looking at just the findings for Hidden Figures, Part 1, and the frequency of analytic questions criterion, one can construct a statistically justified decision rule for subject classification using the Neyman-Johnson technique. In a separately reported analysis of these data, Aiken (1968) obtained results like those pictured in Figure 1. To the left of the vertical division line, video modeling is significantly superior to written modeling, and all subjects in this region should be assigned to video modeling. To the right of the division line, however, the treatments do not differ within the aptitude range. Since video modeling represents a considerably more expensive training treatment than does written modeling, it is appropriate to assign subjects in this region to written modeling, the cheaper treatment.

Combining aptitudes by multiple regression analysis may also clarify the meaning and use of obtained Aptitude × Treatment interactions. In Figure 2, Film Memory and Hidden Figures, which correlate −.10, have been combined to predict frequency of analytic questions on Trial 3 separately for video and written modeling. Note that, with two aptitudes, Aptitude × Treatment interaction is depicted as an intersection of regression planes. The intersection of planes for the two treatments forms a line which, when projected into the base of the figure, divides the bivariate aptitude distribution identifying subjects best assigned to each treatment. It is clear that the use of Hidden Figures and Film Memory scores in combination improves the classification system over what would be obtained using one or the other aptitude variable in isolation. Again, Aiken's (1968) analysis showed video modeling superior to written modeling only for those scoring in the high Film Memory–low Hidden Figures region. Elsewhere the two treatments did not differ significantly.

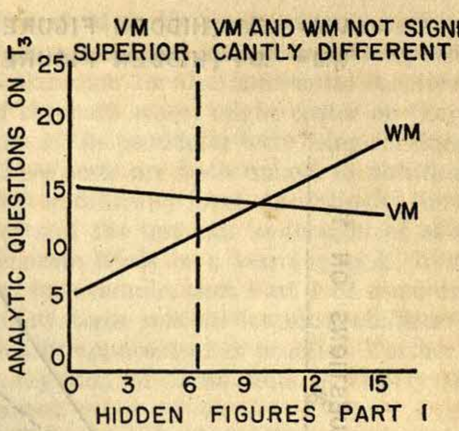


FIG. 1. Regions of significance for a selected Aptitude × Treatment interaction (after Aiken, 1968).

DISCUSSION

The major purpose of this experiment was to examine the effects of trainee aptitudes on observational learning. While Aptitude × Treatment interactions were obtained, their direction did not consistently correspond to predictions. It should be increasingly apparent that hypotheses concerning the direction of Aptitude × Treatment interactions cannot be based only upon superficial content similarities between aptitude tests and learning tasks (Cronbach & Snow, 1969). The results obtained indicate that visual or verbal modes of instruction may or may not be related to corresponding scores on perceptual or verbal aptitude tests. The present crude attempt at task analysis to go beyond content similarities has also had only limited success. However, the obtained Aptitude × Treatment interactions can serve as a basis for further speculation to provide guides for improved Aptitude × Treatment interaction research.

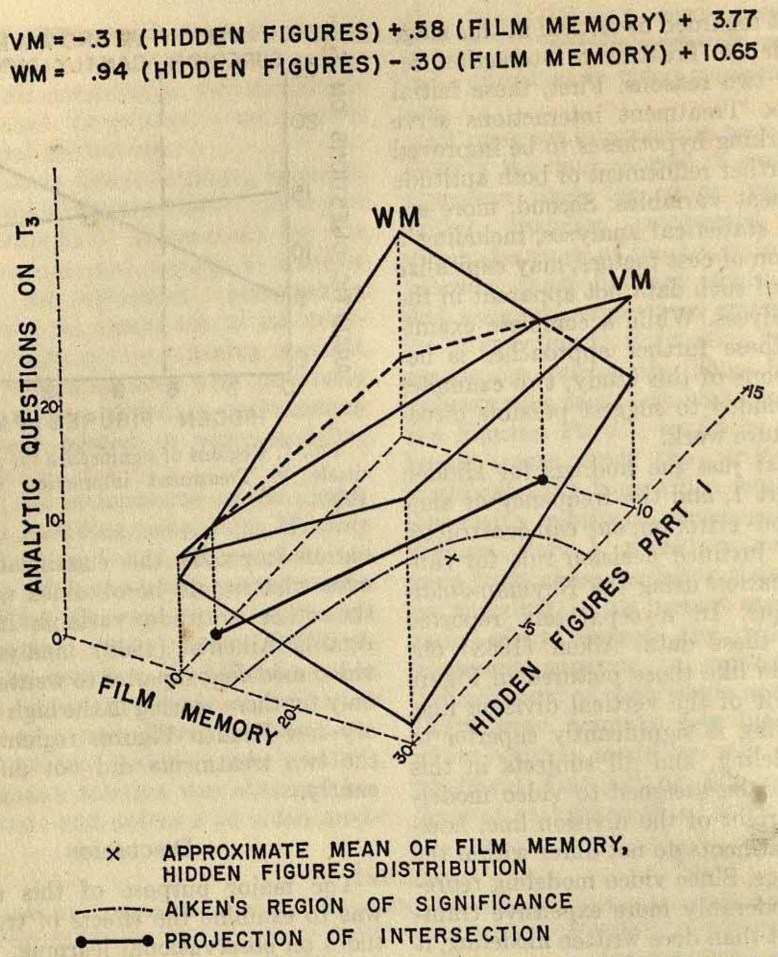


Fig. 2. Multiple regression analysis for video-modeling (VM) and written verbal modeling (WM) treatments.

While it would seem that audiovisual presentations should place demands on perceptual coding abilities, such presentation could also conceivably serve a compensatory function, providing audiovisual representation that might otherwise have to be provided by subjects. The absence of audiovisual content in an instructional presentation may require subjects to generate their own images, thus constituting a demand on perceptual abilities. The direction of obtained Aptitude × Treatment interactions suggests that some abilities may serve such a compensatory function, while others may fulfill a conciliatory or preferential function. Thus, the video-modeling treatment may compensate for deficiencies in some perceptual processing or analytic skills (represented by Hidden Figures) through explicit, concrete presentation of stimulus elements. Such treatment may provide a behavioral representation for the learner that he could not generate for himself if given the written-modeling treatment, and thereby it facilitates his performance. Conversely, for high perceptual-analytic subjects, the presentation of such detailed concrete information at a fixed, perhaps slow pace, may not only fail to be facilitative, but may interfere with encoding processes or attenuate performance through boredom or fatigue. An interesting observation was that learners with high scores on Hidden Figures, Part 1, and Maze Tracing, Part 2, actually seemed to be hindered in the video-modeling treatment, as demonstrated by the negative relationship between aptitude and performance. This was particularly so for those with low Film Memory scores (see Figure 2). For those with higher Film Memory scores, especially when coupled with low Hidden Figures scores, the video-modeling treatment may play to strength. It provides opportunity for preferred visual memory abilities to be exercised.

In the written-modeling treatment, where audiovisual content is not provided, high perceptual-analytic ability subjects appear to do well, particularly if they are also low on Film Memory. Perhaps, following their own unrestricted pace, they can isolate and use key stimuli in the transcript, reviewing for memory as needed. In contrast, subjects low in perceptual-analytic ability may get lost in a verbal maze with the written-modeling treatment and subsequently perform less well.

While Hidden Figures, Part 1, and Maze Tracing, Part 2, were the only variables in the stimulus differentiation component of the model in which significant Aptitude × Treatment interactions were obtained, patterns of correlations for the other ability variables in this component of the model are not inconsistent with this interpretation. Conversely, based upon results obtained using Film Memory scores (as well as in supplementary analyses reported by McDonald & Koran, 1969), the relationship of test to task in the response integration component of the model appears direct rather than compensatory, with scores on Film and Audio-Memory tests positively related to performance in the audiovisual mode of presentation. The exception to this interpretation was the relationship between Film Memory scores and performance on the identification test, but this latter test was a written criterion, quite different in nature from the classroom performance criteria.

Whether Hidden Figures and other such tests are to be regarded as representatives of general fluid ability or of special abilities or styles is uncertain at this point, especially since only part scores for Hidden Figures, Maze Tracing and Sentence Reproduction produced interactions. One possible explanation for the differential functioning of the part scores might center on the nature of the particular tests being considered. These tests are both unique in nature and high in difficulty level. Accordingly, the two parts of the test can be thought of as two separate trials in a learning task. It may be, for example, that Part 1 of some unfamiliar tests reflects learning abilities not clearly apparent after practice. Further investigation of these tests is clearly warranted.

This discussion leads toward a conception of interactions arising from a compensatory-conciliatory process in which some harmony is achieved between learner proficiencies, learner deficiencies, and the characteristics of the treatment. It has been suggested that the theoretical model proposed by Melton (1967) might be used as a plausible basis for hypotheses concerning these kinds of individual differences in learning. Whether there is value in this conception as a guide for identifying Aptitude × Treatment interactions remains a suitable topic for future investigation.

The results of this study support the initial premise that the rate and level of learning of a specific teaching strategy varies as a function of model presentation, and that the effectiveness of instructional methods varies from subject to subject, with such differences being related to trainee aptitudes. Although it has not been demonstrated that the training procedures used represent the most effective way to train teachers, these findings provide evidence that through observation, trainees can acquire principles exemplified in a model's behavior and use them for generating novel combinations of teaching behavior. This research does not lead to detailed suggestions for specific modifications and developments regarding the individualization of teacher training programs. We have only begun to explore the wide range of problems concerned with finding effective training techniques for teachers with different characteristics. However, the results obtained in this experiment appear sufficiently encouraging to suggest the potential value of further research on Aptitude × Treatment interaction in teacher training.

REFERENCES

AIKEN, L. R., JR. *Interactions among group regressions: An old method in a new setting.* (Research and Development Memorandum No. 42) Stanford, Calif.: Stanford Center for Research and Development in Teaching, 1968.

BANDURA, A., & MISCHEL, W. Modification of self imposed delay of reward through exposure to live and symbolic models. *Journal of Personality and Social Psychology*, 1965, **2**, 698–705.

BANDURA, A., & WALTERS, R. Social learning and personality development. Chicago: Holt, Rinehart & Winston, 1963.

BLOOM, B. S. (Ed.), *Taxonomy of educational objectives.* New York: Longmans, Green, 1956.

CRONBACH, L. J., & SNOW, R. E. *Individual differences in learning ability as a function of instructional variables.* (Final Report United States Office of Education OEC 4-6-061269-1217) Stanford, Calif.: Stanford University, 1969.

FRENCH, J. W., ECKSTROM, R. B., & PRICE, L. A. *Kit of Reference Tests for Cognitive Factors.* Princeton, N. J.: Educational Testing Service 1963.

KORAN, M. L. The effects of individual differences on observational learning in the acquisition of a teaching skill. Unpublished doctoral dissertation, Stanford University, 1969.

McDONALD, F. J., & ALLEN, D. W. *Training effects of feedback and modeling procedures on teaching performance.* Stanford, Calif.: Stanford University, School of Education, 1967.

McDONALD, F. J., & KORAN, M. L. The effects of individual differences on observational learning in the acquisition of a teaching skill. (Final Report, United States Office of Education, OEG-9-8-080073-0117(010)) Stanford, Calif.: Stanford University School of Education, 1969.

MELTON, A. W. Individual differences and theoretical process variables: General comments on the conference. In R. Gagné (Ed.), *Learning and individual differences.* Columbus Ohio: Charles E. Merril, 1967.

SEIBERT, W. F., REID, C., & SNOW, R. E. *Studies in cine-psychometry II.* (Final Report, United States Office of Education, 7-24-0880-257) Lafayette, Ind.: Purdue University, Audio-Visual Center, 1967.

SEIBERT, W. F., & SNOW, R. E. *Studies in cine-psychometry I.* (Final Report, United States Office of Education, 7-12-0280-184) Lafayette, Ind.: Purdue University, Audio-Visual Center, 1965.

SKINNER, B. F. *The technology of teaching.* New York: Appleton, Century & Crofts, 1968.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received March 10, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 3, 229-234

INFLUENCE OF RESPONSE MODE, SEX, READING ABILITY, AND LEVEL OF DIFFICULTY ON FOUR MEASURES OF RECALL OF MEANINGFUL WRITTEN MATERIAL[1]

WILLIAM B. TODD[2] AND CLEMM C. KESSLER, III[3]
University of Nebraska at Omaha

Research into the effects of overt versus covert responding in a learning situation has produced contradictory results. The present study was undertaken to clarify some of these findings. Three response modes were compared across three levels of difficulty for men and women with high and low reading ability—a total of 180 college students classified in 36 groups. Comparisons were made on four dependent measures of recall. Consistent differences were found as a function of reading ability; sex was also shown to have significant effects. Few differences were found for difficulty level or response mode. The results demonstrated that (*a*) reading ability must be controlled in studies of learning involving meaningful verbal material; (*b*) an unknown sex factor influences learning or recall of meaningful material; (*c*) the recall of ideas from meaningful material is influenced by the amount of material; (*d*) the facilitating influence of overt responding is limited by the independent and dependent variables involved in the learning or performance situation.

The rapid growth and development of the fields of programmed and computer-assisted instruction has led to the exploration of many variables which are directly relevant to these techniques as well as being generally applicable to basic learning processes. One factor which has received attention has been overt versus covert responding in the learning situation. Skinner (1958) states that an active response facilitates learning and his contention is supported by many studies (see Kemp & Holland, 1966; Price, 1968) and refuted by others (see Crist, 1966; Elliot & MacMurray, 1966).

Previous research has demonstrated that results obtained in studies of overt and covert responding are dependent in part on the following factors: (*a*) the level of difficulty of the material to be learned (Goldbeck & Briggs, 1962; Goldbeck & Campbell, 1962); (*b*) the nature of the material to be learned (Eigen & Margulies, 1963; Goldbeck & Campbell, 1962; Williams, 1966); (*c*) variations in the method of presentation of the material to be learned (Horowitz & Berkowitz, 1967; King & Madill, 1968); (*d*) quantitative or qualitative differences in the dependent measure (Ushkow, 1967; Williams, 1966).

The diverse effects of these variables have prevented the drawing of a general conclusion regarding the desirability of requiring an overt or a covert response in a learning situation. The present study was designed to integrate and systematically explore several of these variables (difficulty, type of recall), to investigate two popular types of overt response which have not been studied to any great extent (underlining and note taking), and to evaluate the influence of two factors which are often left uncontrolled in this area of research, sex and reading ability.

Three response modes were compared: reading (covert response), underlining, and note taking (overt responses) across three levels of difficulty (three lengths of a story) for males and females with high and low reading ability. Comparisons were made on the following dependent measures discussed by King (1960, 1961, 1968) for scoring recalls: (*a*) the total number of words written on the recall regardless of spelling or accuracy; (*b*) the total number of eight-word sequences recalled from the story; (*c*) the total number of idea units in the material recalled (word-for-word similarity and proper sequence were not necessary, merely the essentials of the idea); (*d*) the total number of identical words recalled (a word had to be given as it was in the original passage except for minor spelling or sequence errors. The number of times a repeated word was counted was limited by the number of times it occurred in the original passage).

---

[1] This paper is a summary of a thesis submitted in partial fulfillment of the master's degree in the Department of Psychology, University of Nebraska at Omaha.

[2] Now with the Special Education District, Lake County, Illinois.

[3] Requests for reprints should be sent to the Psychology Department, University of Nebraska at Omaha, P.O. Box 688, Omaha, Nebraska 68101.

The following hypotheses were developed to explore the problem: (*a*) There will be significant differences on each of the dependent measures in terms of level of difficulty, the lowest levels of difficulty resulting in a better recall of each. (*b*) There will be no significant differences between the dependent measures in terms of sex. (*c*) Subjects with high reading ability will recall more on each of the dependent measures than subjects with low reading ability. (*d*) There will be significant differences on each of the dependent measures as a function of response mode.

METHOD

Subjects

The design required 36 groups (3 response modes × 3 levels of difficulty × male and female × high and low reading ability). Each group contained five Introductory Psychology students from the University of Nebraska at Omaha for a total of 180 subjects. Age and class standing were controlled by selecting only subjects under 21 years of age who were admitted as freshmen the previous fall.

Reading ability was determined by using total scores on the Ohio State University Psychological Test. Forty-five of the top and bottom scorers for both males and females were chosen as subjects. There were no overlapping scores for either males or females, a 25-point difference existing between the top low male and the bottom high male and a 17-point difference between the females. The 180 subjects stratified by sex and reading ability were randomly assigned to the nine groups defined by the difficulty and response mode categories.

Materials

The stimulus material consisted of a short story by Bartlett (1932) entitled "War of the Ghosts." There were three levels of difficulty operationally defined as three different story lengths (44 words, 140 words, 256 words). While proportionate lengths of exposure time were allotted to the three passages, it was assumed that even if exposure time and story length were proportionally equivalent, a positive correlation still existed with difficulty; that is, that the learning of a long passage of material is not necessarily of equal ease to that of a short passage (even with proportional times) because of the inherent content differences. Learning difficulty is not just a function of the number of ideas, words, etc., within a period of time. As the Gestaltists would say, the whole is more than the sum of its parts; consequently learning a longer passage even with longer exposure time could be more difficult.

Minor changes were made in the passages to equate them for complexity by employing Flesch's Reading-Ease formula, a technique which has been validated by Chall (1954).

Procedure

A pilot study was conducted to determine the appropriate time limits for presentation and recall of each of the three story lengths. In the pilot study, measures were taken of the time required for 90% of the subjects in the 256-word note-taking group to complete the reading and recall of tasks. Using this measure as a standard, the appropriate proportion of time was then calculated for the other groups. The 140-word subjects received about 55% of the time permitted the 256-word group and the 44-word subjects about 16% of the time for presentation and recall.

After these limits were established, the main experiment was conducted by assembling the subjects according to the response mode–difficulty category to which they were assigned, reading the instructions and collecting the data. The subjects in the reading groups were told to continue reading over the story until time was called. The subjects in the underline groups were told to read the story and underline the important words and phrases but not the whole passage or whole sentences unless they felt them especially important. In the event that they had time to spare after doing this, they were told to read over the underlinings first and then to re-read the story. The subjects in the note-taking groups were told to take notes on a blank sheet of paper which was provided. They were told that outlining was an excellent note-taking method but that they could do it any way they wished. They were instructed to write down important words and phrases but not whole sentences unless they were very important. If they finished before time was called, they were supposed to read their notes again and then read over the story until the time was up.

Scoring

The total number of words, the total number of eight-word sequences, and the total number of identical words were scored by computer. Each story was reduced to idea units and these were scored manually. An interscorer reliability for idea units of .99 was obtained using 25 randomly selected protocols.

RESULTS

The hypotheses were tested using a four-factor analysis of variance. Differences between levels were explored using Duncan's multiple-range test (Bruning & Kintz, 1968). The homogeneity of variance was tested using Cochran's C statistic (Winer, 1962) for each of the four dependent measures. The C values were .0935 for idea units, .0536 for number of words, .0702 for number of eight-word sequences, and .0554 for number of identical words ($C_{.95} = .1400$, $df = 4/36$). The homogeneity of variance assumption was met. Idea units were transformed to percent of the total number possible and the scores of the other three dependent measures were converted to T scores since the raw scores were not comparable across difficulty levels. The level of significance was set at .05 for all tests.

The results for the total number of words recalled are presented in Table 1. There were no significant differences for level of difficulty and response modes but there was a significant difference between males and females (the females recalled more words; $F = 7.44$, $df = 1/144$, $p < .01$), and reading ability (the advanced readers recalled more words than the poor readers; $F = 22.84$, $df = 1/144$, $p < .005$).

The results pertaining to the total number of idea units recalled are presented in Table 2. There was no significant difference for response modes. There was a significant difference for difficulty levels ($F = 31.75$, $df = 2/144$, $p < .005$). The subjects with a middle level of difficulty recalled more idea units than those with the highest difficulty level (minimum critical mean difference = 5.08 $df = 144$, $p < .05$), but the difference between the lowest and middle levels was not significant. The females recalled a greater percentage than did the males ($F = 5.96$, $df = 1/144$, $p < .025$), and the advanced reading ability subjects recalled a larger percentage than did the low-ability groups ($F = 28.32$, $df = 1/144$, $p < .005$).

TABLE 1
GROUP MEANS FOR TOTAL NUMBER OF WORDS RECALLED

Variable	Mean[a]
Difficulty level	
44 words	50.02
140 words	49.95
256 words	50.08
Response mode	
Reading	50.88
Underlining	50.27
Note taking	48.90
Sex	
Male	48.07
Female	51.97
Reading ability	
Low	46.60
High	53.43

[a] Converted to T scores.

TABLE 2
GROUP MEANS FOR TOTAL NUMBER OF IDEA UNITS RECALLED

Variable	Mean[a]
Difficulty level	
44 words	68.13
140 words	65.80
256 words	49.91
Response mode	
Reading	60.88
Underlining	64.21
Note taking	58.75
Sex	
Male	58.80
Female	63.76
Reading ability	
Low	55.88
High	66.68

[a] Converted to percentages.

Table 3 presents the data for the total number of eight-word sequences recalled. There were no significant differences for sex or difficulty levels. There was a difference for reading ability, the high-ability groups recalling more sequences than the low-ability groups ($F = 36.88$, $df = 1/144$, $p < .005$). There was also a difference in response modes ($F = 11.83$, $df = 2/144$, $p < .005$), the subjects taking notes having the poorest recall, the reading groups next, and the underlining groups producing the greatest number of recalls (minimum critical mean difference = 3.12, $df = 144$, $p < .05$).

The results for the total number of identical words recalled are displayed in Table 4. There were no significant differences in terms of difficulty level or response modes. The females recalled more identical words

TABLE 3

GROUP MEANS FOR EIGHT-WORD SEQUENCES RECALLED

Variable	Mean[a]
Difficulty level	
44 words	50.42
140 words	50.05
256 words	50.28
Response mode	
Reading	50.15
Underlining	53.97
Note taking	46.63
Sex	
Male	49.92
Female	50.58
Reading ability	
Low	46.51
High	53.99

[a] Converted to *T* scores.

than the males ($F = 4.72$, $df = 1/144$, $p < .05$) and the subjects in the high-ability group recalled more identical words than did those in the low-ability group ($F = 34.34$, $df = 1/144$, $p < .005$).

There were no significant interactions of the independent variables for any of the four dependent measures. Table 5 presents a summary of the findings for the four independent and four dependent variables.

Intercorrelations were run between the four dependent measures. The total number of words, the total number of idea units, and the total number of identical words recalled correlated .927 or greater with each other. The total number of eight-word sequences recalled had low correlations with the other variables, the highest being .111.

DISCUSSION

It was hypothesized that there would be a significant difference in the performance of subjects with high and low reading ability on each of the dependent measures. This hypothesis was supported. Good readers would be expected to be more familiar with words and their usage, have a better vocabulary, and be able to read faster. This was reflected in the total number of words and the total number of identical words recalled. The idea units and eight-word sequences supposedly represented organizational measures and, again, it was anticipated that better readers would prove superior at organization and comprehension.

A sex difference was found for each of the dependent measures except for the recall of eight-word sequences. Thus, the hypothesis of no sex differences received little support. This finding contradicts previous research which indicated that there would be no sex differences in reading ability at the upper grade levels (King, 1960; King & Madill, 1968; King & Russell, 1966; Wozencraft, 1967). An ex post facto inspection of the Ohio State reading scores revealed that there were no significant differences between sexes in either the high- or low-level groups: the sex groups were matched with respect to reading ability.

The disagreement with previous research may be due in part to variations in method. In the current study, the subjects were not limited to one exposure to the material: if they finished before the time limit, they were permitted to review it. In addition, time limits were placed on the recalls. This may have resulted in different performance than would have been obtained with untimed recalls. Regardless, there appeared to be a strong influence on the three dependent measures found to be correlated. This would suggest that sex-related differences exist which must be controlled in future research and attended to in the learning situation. The variables which might account for this

TABLE 4

GROUP MEANS FOR TOTAL NUMBER OF IDENTICAL WORDS RECALLED

Variable	Mean[a]
Difficulty level	
44 words	49.97
140 words	50.00
256 words	50.03
Response mode	
Reading	50.35
Underlining	51.57
Note taking	48.08
Sex	
Male	48.50
Female	51.50
Reading ability	
Low	45.96
High	54.04

[a] Converted to *T* scores.

TABLE 5

SUMMARY OF FINDINGS FOR THE FOUR DEPENDENT MEASURES AS A FUNCTION OF THE FOUR INDEPENDENT VARIABLES

Independent variables	Dependent measures			
	Total words recalled	Idea units recalled	Eight-word sequences recalled	Identical words recalled
Difficulty level	No differences	Subjects exposed to short and medium length stories recalled more than subjects exposed to long story	No differences	No differences
Response mode	No differences	No differences	Underlining produced the highest number of recalls, then reading, then note taking.	No differences
Sex	Females recalled more than males	Females recalled more than males	No difference	Females recalled more than males
Reading ability	Good readers recalled more than poor readers	Good readers recalled more than poor readers	Good readers recalled more than poor readers	Good readers recalled more than poor readers

difference range from IQ to motivational pattern.

It was predicted that there would be differences in recall as a function of level of difficulty of the material, the shortest story producing the greatest number of recalls for each dependent measure, and the longest story producing the fewest. This hypothesis was confirmed only for the idea units dependent measure. In this case the short and medium length stories were recalled better than the long story. It appears that increasing story length may inhibit comprehension but have less influence on the more rote aspects of learning or retention.

The variable of primary interest, response modes, produced significant differences only on the eight-word sequences where underlining was found to be most effective. Differences were expected on each of the dependent measures as had been found in studies where writing was used as the overt response (Cummings & Goldstein, 1964; Ushkow, 1967; Williams, 1966). It should be noted that the eight-word sequence was a unique measure in that it did not correlate highly with the other three dependent variables. Also, the eight-word sequence was the one measure where sex differences did not appear. Although it is difficult to explain these differences without further research, at least two possibilities exist. One of these involves the nature of the material (the passage in the present study was meaningful, in other studies it consisted of more discrete and factual material), the other being the procedure used in that previous studies permitted one exposure to the material whereas the subjects could have reviewed it several times in the current study.

One conclusion seems inescapable, however. Overt responding may be facilitating, but only with certain types of material, under particular conditions, and only when certain measures of recall are used. This finding has also been suggested by other studies and would indicate that the generalizability of the overt responding concept is severely limited and must be subjected to further investigation in order to determine the learning and performance conditions under which it is most efficacious.

REFERENCES

BARTLETT, F. C. *Remembering: A study in experimental and social psychology*. London: Cambridge University Press, 1932.

BRUNING, J. L., & KINTZ, B. L. *Computational handbook of statistics*. Glenview, Ill.: Scott, Foresman, 1968.

CHALL, J. S. Readability: An appraisal of research and application. *Bureau of Educational Research Monographs*, 1954, No. 34.

CRIST, R. L. Overt vs. covert responding and retention by sixth-grade students. *Journal of Educational Psychology*, 1966, **57**, 99–101.

CUMMINGS, A., & GOLDSTEIN, L. S. The effect of overt and covert responding on two kinds of learning tasks. In J. P. DeCecco, *Educational technology: Readings in programmed instruction*. New York: Holt, Rinehart & Winston, 1964.

Eigen, L. D., & Margulies, S. Response characteristics as a function of information level. *Journal of Programed Instruction,* 1963, **2,** 45–54.

Elliott, R., & MacMurray, T. Effectiveness of two methods of presenting information about simple reflexes. *Psychological Record,* 1966, **16,** 283–288.

Goldbeck, R. A., & Briggs, L. J. An analysis of response mode and feedback factors in automated instruction. (Technical Report No. 2) Santa Barbara: American Institute for Research, 1960. Cited in R. A. Goldbeck & V. N. Campbell, The effects of response mode and response difficulty on programed learning. *Journal of Educational Psychology,* 1962, **53,** 110.

Goldbeck, R. A., & Campbell, V. N. The effects of response mode and response difficulty on programed learning. *Journal of Educational Psychology,* 1962, **53,** 110–118.

Horowitz, M. W., & Berkowitz, A. Listening and reading, speaking and writing: An experimental investigation of differential acquisition and reproduction of memory. *Perceptual and Motor Skills,* 1967, **24,** 207–215.

Kemp, F. D., & Holland, J. G. Blackout ratio and overt responses in programed instructions: Resolution of disparate results. *Journal of Educational Psychology,* 1966, **57,** 109–114.

King, D. J. On the accuracy of written recall: A scaling and factor analytic study. *Psychological Record,* 1960, **10,** 113–122.

King, D. J. Scaling the accuracy of recall of stories in the absence of objective criteria. *Psychological Record,* 1961, **11,** 87–90.

King, D. J. Retention of connected meaningful material as a function of modes of presentation and recall. *Journal of Experimental Psychology,* 1968, **77,** 676–683.

King, D. J., & Madill, J. Complex methods of presentation, internal consistency of learning material, and accuracy of written recall. *Psychological Reports,* 1968, **22,** 777–782.

King, D. J., & Russell, G. W. A comparison of rote and meaningful learning of connected meaningful material. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5,** 478–483.

Price, A. H. The effects of subject response mode and peer social reinforcement of childrens learning in programmed instruction. *Dissertation Abstracts,* 1968, **28**(1-A), 131–132.

Skinner, B. F. Teaching machines. *Science,* 1958, **128,** 969–977.

Ushkow, E. A. The influence of criterion test type and the amount of delay between learning and testing on the effectiveness of overt and non overt responding to programmed instruction. *Dissertation Abstracts,* 1967, **28**(4-B), 1716–1717.

Williams, J. P. Combining response modes in programmed instruction. *Journal of Educational Psychology,* 1966, **57,** 215–219.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

Wozencraft, M. A comparison of the reading abilities of boys and girls at two grade levels. *Journal of the Reading Specialist,* 1967, **6,** 136–139.

(Received March 10, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 3, 235–239

INSTRUCTOR ATTITUDES TOWARD, AND THEIR USE OF, STUDENT RATINGS OF TEACHERS[1]

MARTIN T. MILLER[2]
University of Iowa

Thirty-six teaching assistants in three freshman courses were assigned to groups on the basis of an attitude scale measure regarding the value they ascribe to student ratings of teachers. Then they were randomly assigned to one of two groups, one of which received feedback from students' ratings during the semester. Differences between attitude and feedback groups were examined using as criterion measures (*a*) end of semester ratings and (*b*) final examination scores, adjusted by analysis of covariance using midsemester ratings and examination scores as covariates. Instructors in feedback or attitude groups did not differ significantly on their end-of-semester ratings. In two of the classes there were no significant differences on the final exam scores for the feedback and attitude groups. In the third class, final examination scores were higher for those students whose teachers received feedback from the ratings.

The solicitation of student ratings of teachers can be justified by an observation by Remmers (1958) that

> No teacher has any choice as to whether he wishes to be judged by his students. The only choice he has is whether he wishes to know how he is judged and thus possibly capitalize on this feedback [p. 20].

Daw and Gage (1967) hypothesized that feedback to teachers operated to induce a state of disequilibrium resulting from a discrepancy between the self-concept of the person being rated and his perception of how he is regarded by the rater. Tuckman and Oliver (1968) found that high school teachers who received feedback from student ratings received improved subsequent ratings significantly more than those who did not receive feedback. (Improved here means that ratings changed in the direction suggested by the students.) A second finding was that teachers' response to feedback from principals also resulted in a change in behavior, but in this instance in the direction opposite to the suggestions made by the principals. Thus, it would appear that the instructor attitude toward the source of suggestions has an effect on the response the teachers make to the information provided by such feedback.

While ratings of instructor behavior may provide one source or type of information to the teacher, Mitzel and Gross (1958) stated that teachers may be considered effective to the extent that they stimulate their pupils to achieve desired educational ends. Among such goals no doubt would be the academic growth of these pupils. However, in their reviews of research concerned with correlations between teachers' ratings and pupil attainment, Stecklein (1960) and Medley and Mitzel (1963) concluded there was no appreciable relationship. Thus, it is not surprising that Gustad (1967) found a decline in the rank order of student ratings of teachers, as teacher evaluation sources, between 1961 and 1966. Gustad interpreted this decline to be due at least in part to a lack of empirical validity associated with student ratings.

The general conclusion of an absence of a significant relationship between teacher ratings and student academic growth has been criticized for the frequent lack of control in both the design and analyses of the data. As an example, Mitzel and Gross (1958) considered it "illogical" to combine tests in different subject areas, despite standard score conversions, when attempting to measure a unitary teacher effectiveness dimension. They also raised doubts on the appropriateness of the statistical methods used in many of the studies of pupil growth. Similar objections have been raised by Gage (1958) who urged the use of instructors of similar background, teaching under uniform conditions, students in introductory multisectioned courses. Therefore an attempt was made in this study to control relevant characteristics of students and instructors as well as course content by both experimental as well as statistical procedures.

---

[1] This study is based on the author's doctoral dissertation at the University of Iowa. Appreciation is extended to Dale P. Scannell for his valuable guidance in the development of the study.

[2] Requests for reprints should be sent to Martin T. Miller who is now at the Psychology Department, Bard College, Annandale-on-Hudson, New York, New York 12504.

The purpose of the present study was to determine empirically whether providing instructors with information from student ratings had effects on their subsequent ratings by students and on student achievement. In addition, the study was to assess whether these effects were a function of instructor attitudes toward the value of the student ratings.

The criterion measures used in assessing the effect of feedback from student ratings were (*a*) a measure of class academic achievement, and (*b*) ratings at the end of the semester. Academic achievement, as measured by departmental examinations, was assumed to provide a measure of the product of the teacher's efforts and skills. The student ratings of instruction were assumed to provide a measure of some aspects of the teaching process. The relationship between these product and process criteria was investigated to determine whether feedback on the process measure (student ratings) affected subsequent ratings, or had an effect on the product criterion (student academic achievement).

METHOD

Subjects

Thirty-six graduate teaching assistants in Religion ($N = 14$), and two courses in Earth Science (Ns = 10 and 12) were assigned at random to discussion sections in their respective courses. Each teaching assistant was assigned to at least two, and not more than three, discussion groups of approximately 20 students.

Approximately 2,000 students enrolled in these undergraduate freshman level courses were assigned at random to the discussion sections within each course.

In all instances the teaching assistant was considered the unit of analysis. This means that data were combined for all sections taught by each teaching assistant.

Instruments

The Survey of Student Opinion of Teaching (SSOT) used in this study was a revision of a scale constructed by Stuit and Ebel (1952). The revision prepared in 1967 was comprised of 15 items, 10 of which referred to instructor characteristics and 5 to characteristics of the course. Norms for each teaching rank were derived in 1968 and included data for 391 classes with over 10,000 students taught by teaching assistants.

The questionnaire on instructor attitudes toward the value of student ratings developed for this study was based on the first 10 items of the SSOT, those reflecting instructor characteristics. The teaching assistants were told the purpose of the questionnaire was "to assess the value instructors ascribe to student ratings as a source of suggestions which may be used to modify teacher behavior." Teachers were instructed to rate independently, on a 5-point scale, the value to them of a student evaluation of each of the teacher characteristics.

Student academic achievement was measured by tests constructed from items submitted and discussed by the teaching assistants. It was assumed that these items more clearly reflect the common goals and content than would be the case were an outside test source used. All examinations were administered on a departmental basis and scored and analyzed by the university examination service.

Procedure

The instructor attitude questionnaire was administered to the teaching assistants during the fourth week of the semester. Median scores on the attitude measure were obtained separately for the teaching assistants in each of the three courses and provided the basis for establishing the groups with "favorable" and "unfavorable" attitudes toward the student ratings. Each of these groups was divided at random into two subgroups, one of which was to receive feedback from student ratings early in the semester. Thus, four treatment groups were established for teachers in each of the three courses: (*a*) favorable attitude, received feedback (*b*) unfavorable attitude, received feedback (*c*) favorable attitude, did not receive feedback (*d*) unfavorable attitude, did not receive feedback

The SSOT forms were administered and scored for all teachers, and returned to the selected instructors during the fifth week of the semester prior to the administration of the first examination. At the end of the semester all students again rated their instructor prior to the administration of the final examination.

The results of the first midsemester examination and the final examination were obtained for the sections taught by the teaching assistants in each of the three courses.

ANALYSIS

Differences among the teaching assistants in the four treatment conditions in each of the three classes were examined by an analysis of covariance. In this way the treatment groups were adjusted statistically for level of midsemester ratings so that differences in final ratings might be related to the treatment effects. Similarly, the treatment groups were adjusted on midsemester examination scores to relate to differences in final examination scores.

RESULTS

The adjusted mean end-of-semester ratings on the SSOT are presented in Table 1. The analyses of covariance to test the significance of the differences for both feedback and attitude groups on the SSOT at the end of the semester in each of the three courses was nonsignificant ($F < 1.0$; $df = 1/12, 1/8,$ and $1/10$).

These findings suggest that the average end-of-semester ratings for teaching assistants who had access to midsemester ratings of their instruction did not differ significantly from the average ratings assigned to teaching assistants who did not have access to the ratings. Further, there were nonsignificant differences between instructor final ratings as a function of their attitudes toward the value of the ratings. That is, the mean ratings for the favorable and unfavorable groups did not differ significantly.

TABLE 1

ADJUSTED MEAN END-OF-SEMESTER RATINGS ON THE SURVEY OF STUDENT OPINION OF TEACHING FOR FEEDBACK AND ATTITUDE GROUPS

Course	Attitudes		
	Favorable	Unfavorable	Total
1			
Feedback	3·93	3·84	3·88
No feedback	3·85	3·99	3·92
Total	3·89	3·91	
2			
Feedback	3.32	3·57	3·45
No feedback	3.43	3·93	3·68
Total	3.36	3.71	
3			
Feedback	3.74	3.76	3.75
No feedback	3.78	3.65	3.72
Total	3.76	3.71	

TABLE 2

ADJUSTED MEAN END-OF-SEMESTER EXAMINATION SCORES FOR FEEDBACK AND ATTITUDE GROUPS

Course	Attitudes		
	Favorable	Unfavorable	Total
1			
Feedback	65.13	64.74	64.94
No feedback	63.37	64.74	64.05
Total	64.38	64.74	
2			
Feedback	63.47	64.99	64.23
No feedback	63.02	69.59	66.30
Total	63.29	66.83	
3			
Feedback	59.49	59.87	59.68
No feedback	56.92	55.78	56.35
Total	58.20	57.83	

The adjusted mean end-of-semester examination scores are presented in Table 2. The analyses of covariance to test the significance of the differences for both feedback and attitude groups on final examination scores showed an inconsistency among the three courses. For Courses 1 and 2, the differences between mean final examination scores for sections taught by teaching assistants in the two feedback conditions was not significant ($F < 1.0$; $df = 1/12, 1/8$) regardless of instructor attitudes toward the value of the ratings. However, for the teaching assistants in Course 3, the results show that the difference between instructors in the feedback conditions was significant ($F = 15.77$, $df = 1/10$, $p < .01$). Further, the mean final examination scores were higher for those students whose instructors received feedback from students' ratings during the semester than for those students whose instructors did not.

Correlations between instructor ratings on the SSOT at the beginning and end of the semester are presented in Table 3. The results show that for Courses 1 and 2 (rs $=$ .76 and .76) the students did not alter their perceptions of their instructors as much as for Course 3 ($r = .38$), the course in which the instructors who received feedback from

TABLE 3

CORRELATIONS BETWEEN MEAN INSTRUCTOR RATINGS ON THE SURVEY OF STUDENT OPINION OF TEACHING AT THE MIDDLE AND END OF THE SEMESTER

Course	r	N	p
1	.76	14	<.01
2	.76	10	<.02
3	.38	12	<.30

the ratings had sections with higher academic achievement.

DISCUSSION

The results of the study suggest that when teachers of similar academic and teaching experience present similar course content to students assigned randomly to discussion sections of the same class, and when the same criterion measures are used and subjected to appropriate statistical analysis, feedback from student ratings does not seem to alter instructor behavior as manifested in subsequent student ratings. Further, ratings do not appear to be changed as a function of the attitude of the teacher toward receiving students' ratings.

A second finding from the study was that, for the instructors in two of the three courses, feedback from students' ratings did not improve instruction and hence the academic performance of the students. Once again instructor attitudes toward the value of the ratings was not a significant factor. However, for the teaching assistants in the third course, there was a significant difference in the academic achievement between those classes whose instructors received feedback from students' ratings and those who did not. Further, the correlation between ratings at the midsemester and the end of the semester showed that the teaching assistants were being perceived somewhat differently. The change in ratings was not however, directly associated with either the feedback or attitude conditions.

The differences between the teaching assistants' uses of the ratings in the three courses are difficult to explain. The students in the discussion sections were randomly distributed. The teachers were similar in classroom experience and in academic background. Teachers were rated on a common instrument and reported their attitudes toward the ratings on a common scale. The departmental examinations were constructed in a similar manner and were objectively scored.

A partial explanation of the difference may include such factors as the motivation of the teachers to use the results of the student ratings which cannot be directly inferred from their attitudes toward the value of the ratings or their abilities to use the feedback. As teaching assistants, many perhaps had not made a commitment to a teaching career and thus their motives for improving their skills were somewhat uncertain. Also, the use of the instructor as the unit of all the analyses may have resulted in sampling errors due to the small N per cell.

While conclusions drawn from the results of the study are limited, the results suggest some limitations in the use of student ratings as a method of improving instruction. Further analyses of data collected in this study are being conducted to determine whether specific items on rating instruments differentially reflect the effects of feedback. Future research may be directed profitably in the investigation of skills or characteristics of instructors who seem able to use the results of student surveys.

REFERENCES

DAW, R. W., & GAGE, N. L. Effect of feedback from teachers to principals. *Journal of Educational Psychology,* 1967, **58,** 181–188.

GAGE, N. L. Ends and means in appraising college teaching. In W. J. McKeachie (Ed.), *The appraisal of teaching in large universities.* Ann Arbor: University of Michigan, 1958.

GUSTAD, J. Evaluation of teaching performance: Issues and possibilities. In C. Lee (Ed.), *Improving college teaching.* Washington, D.C.: American Council on Education, 1967.

MEDLEY, D., & MITZEL, H. Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.

MITZEL, H., & GROSS, C. F. The development of pupil growth criteria in studies of teacher effectiveness. *Educational Research Bulletin,* 1958, **37,** 178–187, 205–215.

REMMERS, H. H. On students' perceptions of teachers' effectiveness. In W. J. McKeachie (Ed.), *The appraisal of teaching in large universities.* Ann Arbor: University of Michigan, 1958.

STECKLEIN, J. E. Colleges and universities. In C. W. Harris (Ed.), *Encyclopedia of educational research.* (3rd ed.) New York: Macmillan, 1960.

STUIT, D. B., & EBEL, R. L. Instructor ratings at a large state university. *College and University* 1952, **27,** 247–254.

TUCKMAN, B. W., & OLIVER, W. F. Effectiveness of feedback to teachers as a function of source. *Journal of Educational Psychology*, 1968, **59,** 297–301.

(Received March 10, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 3, 240–244

EFFECTS OF RACE OF TESTER AND NEED FOR APPROVAL ON CHILDRENS' LEARNING

CHARLES TURNER[1]
Columbia University

Eighty white male children were given a simple serial learning task by experimenters of different racial identities. The experimenters motivated the children with verbal approval during the testing. It was found that those subjects tested by same-race male testers performed significantly better than subjects tested by different-race male experimenters. No correlation was found between subjects need for approval as measured by the Marlowe-Crowne Social Desirability test and the speed of their learning.

There are strong feelings among many psychologists (e.g., Katz, 1964) that the race, social class, and sex of an experimenter have an effect on the motivational level of his subjects, and thus, on the subjects' performance. However, "experimental research in this area is meager and uncertain [Rosenhahn, 1966, p. 253]." As Katz, Henchy, and Allen (1968) have noted, what is needed to develop and test a hypothesis concerning the effect of these factors is a series of experiments which would determine childrens' responses to testers of different racial and class identities. The present experiment is a further step in this direction; it presents a general hypothesis concerning the nature and effect of such factors, and tests the hypothesis using white children, and experimenters of different racial identities.

Bandura (1964) has found that a young child will imitate most readily those adults who are "dispensers of rewards" in his environment. Generalizing from this finding, it would seem plausible that young children would identify with, and be most susceptible to influence from, people who closely resemble the dispensers of reward in the childrens' home environment. Thus as an experimenter became more similar to a young child's important others (dispensers of rewards) in appearance, demeanor, and other relevant dimensions, he would become more effective in motivating and directing the child's behavior.

In the few interracial experiments carried out so far this hypothesis appears to be tenable. Aronson and Golden (1962) have found that white school children respond to the race of a communicator by either raising or lowering their receptivity to the point of view which the communicator espoused. It was found in particular that those children who evidenced more negative feelings toward blacks on a questionnaire tended to show less attitude change with black communicators (versus white communicators) even when the communicator had high prestige in relevant areas. Similarly, those subjects who showed more positive feelings toward blacks evidenced more attitude change when the communicator was black rather than white. For the "prejudiced" children the results fit the hypothesis which was generalized from the work of Bandura, Ross, and Ross (1963). For the "less prejudiced" children the argument can be made that they may have been exposed to more amicable interaction with black adults, perhaps in situations where these black adults were dispensers of primary rewards. It is also possible that there were other overt social factors working differentially for the two groups of children. Such factors as the behaviors and sanctions of "more liberal" parents may have influenced those children who gave less negative opinions about blacks in the questionnaire and who tended to be unusually open to suggestions from black communicators. Thus, besides the influence of some primary prejudice or unfamiliarity variable, the particular environments of the less prejudiced children may have reinforced them with social approval for openness and positive responses to black adults.

---

[1] The author wishes to thank Dan Ehrlich and Richard Christie for their helpful suggestions.

The author is currently at the Department of Social Psychology, London School of Economics and Political Science.

In a more recent experiment Katz, Henchy, and Allen (1968) demonstrated that 7- to 10-year-old black children will learn items in a memory test faster when they are motivated with verbal approval or disapproval by black testers (vis á vis white testers). This finding led Katz et al. (1968) to theorize that the negative feelings which white adults and especially white teachers have toward black youth are perceived by them and cause them to respond less well to verbal approval by whites. This hypothesis does not predict a similar effect for white children when tested by individuals of different racial identities since black teachers do not have similar negative attitudes toward white students (cf. Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, & York, 1966).

In order to test the effect of Katz's paradigm on white children when testers of different racial identities are employed the present experiment was undertaken. It was predicted, in line with the theory extended from the work of Bandura, that white children would respond better to motivation (in the form of verbal approval) from adults who were of the same racial identity as their normal dispenser of rewards, that is, white adults.

The present study used the experimental design formulated by Katz et al. (1968) but with white children serving as subjects. Those subjects tested by different-race (black) testers were expected to have lower learning scores than those subjects tested by same-race (white) testers.

The effect of the childrens' need for approval as reflected by the Marlowe-Crowne Social Desirability scale was also measured. It has previously been shown (Epstein, 1964) that children who scored high on the Marlowe-Crowne scale are more readily conditioned using verbal reinforcement than low scorers. Katz et al. (1968) found this to be an important factor in analyzing the results of such experiments as the present one. So, it was further predicted that those subjects who evidenced a high need for approval would have higher learning scores than subjects with a low need for approval.

METHOD

Subjects and Experimenters[2]

Eighty white boys in the ninth-grade class (ages: 13–15 years) of an all male parochial school served as subjects. The student population of this Catholic school was predominantly (96%) white, and the school's faculty was all male and all white. An analysis of parental occupations using the classification system developed by Warner, Meeker, and Eells (1949) indicated that 94% of the subjects belonged to the three lowest socioeconomic groups.

Testing was done individually by eight experimenters: two black males; two black females; two white males; and two white females. The experimenters were 20–25 years of age, middle-class, and were unaware of the hypothesis of the experiment.

General Procedure

Groups of 10 children were randomly asigned to each experimenter, who tested each child individually. The testing was carried out in private offices at the childrens' school. Each boy was taken individually from his regular classes, and was met at the testing area by the experimenter. After a few minutes of friendly interaction designed to put the subject at ease the experimenter gave the child a 10-item version[3] of the Marlowe-Crowne Social Desirability test to complete.

After the subject had completed the Marlowe-Crowne form the experimenter explained that he wished "to study how children learn things," and that the testing had "no relation to school work and would not be seen by anyone in your school." The experimenter then explained and demonstrated the task which was required of the child. The test was a simple learning task in which a list of eight CVC trigrams of .78 mean association value (from Archer, 1960) were presented on a memory drum. The method of serial anticipation was used, and all subjects were run through nine trials after the initial presentation. Each trigram was automatically exposed for 3 seconds, and there was an intertrial interval of 15 seconds.

Evaluative Feedback

At the beginning of the experiment each subject was encouraged to "try to do your best and learn the syllables as quickly and correctly as possible." Feedback was given three times during the testing. At the end of the first trial the experimenter called the child by name and then said, "You are doing very well, I am very pleased with how well you are doing." With slight alterations this statement was repeated at the end of the third and fifth trials.

---

[2] The author wishes to express his appreciation to the staff and student body of Most Holy Trinity High School for their cooperation in providing subjects.

[3] Consisting of Items 2, 3, 4, 6, 9, 10, 14, 15, 19, and 31 (from Marlowe and Crowne, 1960).

RESULTS

Table 1 shows the mean number of items correctly identified (learning score) by the subjects in nine trials. The data are broken down according to the race and sex of the experimenters, that is, black males, white males, black females, and white females, and the standard deviation and sample size are indicated for each cell. For male testers subjects showed higher mean learning scores with same-race (white) testers. This effect is shown in Figure 1 where the mean learning scores for each third of the experiment are shown for subjects tested by black male and white male testers. A similar effect is not evident in the mean learning scores of subjects tested by female testers. An analysis of variance was applied to the data to discover if the subjects responded differentially to the testers' sexual or racial identity or to an interaction of these two factors. There was a significant effect ($F = 5.6$, $df = 1/76$, $p < .05$) for the manipulation of the racial variable; however the sex variable did not produce a significant result either as a main effect ($F = .1$, $df = 1/76$) or in interaction with the race variable ($F = 0.0$, $df = 1/76$).

The results obtained by individual experimenters within each of the four categories of testers were similar in mean learning score and variance, except for the white female testers. An examination of the subjects randomly assigned to the two white female testers revealed that they were very similar in socioeconomic status and had highly similar mean Marlowe-Crowne Social Desirability scores, however the mean learning scores obtained from the two groups were significantly different ($t = 2.2, p < .05$).

TABLE 1

MEAN LEARNING SCORES FOR FOUR CATEGORIES OF TESTERS

Learning scores	Testers			
	Males		Females	
	Black	White	Black	White
$\bar{X}$	33.50	39.70	39.20	38.45
SD	10.65	12.27	11.10	11.57
n	20	20	20	20

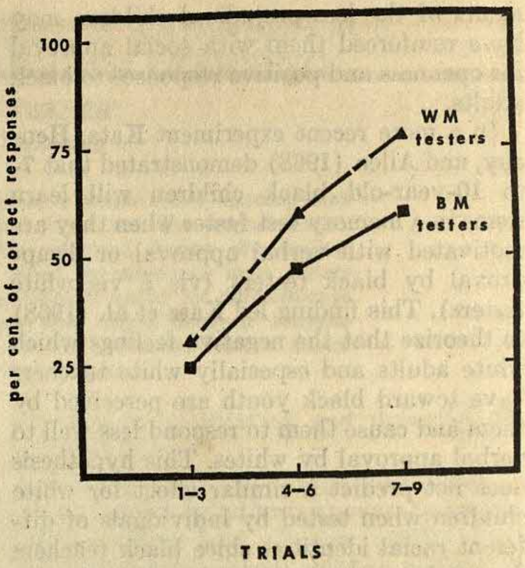


FIG. 1. Mean learning scores for children tested by white and black male testers.

The effect of the subjects' need for approval as measured by the subset of the Marlowe-Crowne Social Desirability test was analyzed by computing Kendall's rank order correlation coefficient for the Marlowe-Crowne scores and the learning scores both for the entire population and for each of the four subgroups. Kendall's test did not show a significant relationship between the Marlowe-Crowne scores and learning scores in any group. For the entire population the correlation coefficient revealed an insignificant negative correlation ($t = -.06$) between the Marlowe-Crowne scores and learning scores. The subset of items used in this test showed similar distributions on a split-half reliability test; $M(\text{even}) = 1.81$, $M(\text{odd}) = 1.76$, $SD(\text{even}) = 1.03$, $SD(\text{odd}) = 1.05$. The reliability coefficient was only .479, which is a reflection, in part, of the small number of items (five) in each half of the test.

DISCUSSION

The results indicate that, at least in situations as limited as the present experiment, persons of the same race as subjects are most effective in motivating them to do rote-memory learning. Katz et al. (1968) have shown this effect with black children and experimenters of different racial identities, and the present experiment has revealed a similar effect with white subjects. This evidence supports the hypothesis that was developed from the Bandura experiments: Children identify with, and are most susceptible to influence from persons who resemble the dispensers of rewards in their normal environment. Since parents are the normal "dispensers of rewards" for their children, it is reasonable that children will be most motivated by testers of the same race as their parents.

In explaining why black male children performed better with black testers Katz et al. (1968) suggested "the notion that disadvantaged black children tend to feel disliked by white middle-class teachers [p. 42]." In support of this explanation they cited attitudinal studies of Northern teachers (Clark, 1965; Gottlieb, 1964) which show that white teachers express a preference for teaching white children. However this hypothesis does not explain why the disadvantaged white children in the present study would perform best with white middle-class testers, since there has been no analogous finding in attitudinal studies of black teachers. Coleman et al. (1966) have reported that none of the black teachers they interviewed expressed a preference for teaching in an all black school. The "dispenser of rewards" hypothesis derived from Bandura's theories does explain this phenomenon, and it seems capable of providing a more adequate and parsimonious explanation of the data. At this point however, the novelty of the interracial situation for the white subjects in the present experiment cannot be overruled as a possible explanation of the diminished performance of the interracial dyads. This "novelty" theory would not however explain the data reported by Katz et al. since his black subjects had already been exposed to the New York City school system where there is an overwhelming majority of white teachers.

The high variability found between white female experimenters is consistent with previous findings and the peculiarities of the subjects' environment. It has been reported (Stevenson, 1965) that there are strong cross-sexual effects in motivational studies, and that different women have highly differential effects on subjects. This report was confirmed in the present study by the significantly different results obtained by the two white female experimenters who tested highly similar groups of children. The cross-sexual effect cited by Stevenson (1964) was probably heightened in this study by the peculiarities of the boys' environment, that is, an all male high school with an all male staff. This would serve to increase the variability of results obtained with female testers and is a possible explanation for the similar mean learning scores obtained by white female and black female testers.

It was also found, using all 80 children, that there was no significant correlation between Marlowe-Crowne Social Desirability scores and the children's learning scores. A detailed analysis of the data did not reveal a significant correlation between these scores and learning scores for any single experimenter category or combination of experimenter categories. This finding is not consistent with the results obtained by Katz et al. (1968) with black children, or with the results obtained by Epstein (1964) with asthmatic grade school children. Although the use of a 10-item subset of the questionnaire is not an optimal indicator of the predictive usefulness of the test, the distributions of scores for each split-half of the subtest were highly similar and the .479 reliability coefficient for such a short questionnaire suggest that there was some degree of internal consistency among the items chosen. In addition Katz et al. have made similar use of a 7-item subset of the childrens' version (cf. Crandall, Crandall, & Katkovsky, 1965) of the Marlowe-Crowne Social Desirability test in their experimental research which corroborated the link between need for approval and the effectiveness of verbal reinforcement in motivating learning. The lack of significant correlation between these two variables in the present experiment suggests the need for further study before accepting the Marlowe-Crowne scores as predictors of childrens' responses to verbal approval; it additionally points up the need for administering entire tests in future research in this field.

REFERENCES

Archer, E. J. Reevaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74**(10, Whole No. 497).

Aronson, E., & **Golden**, B. The effect of relevant and irrelevant aspects of communicator credibility on opinion change. *Journal of Personality*, 1962, **30**, 135–146.

Bandura, A., Ross, D., & Ross, S. A comparative test of the status envy, social power, and secondary reinforcement theories of identificatory learning. *Journal of Abnormal and Social Psychology*, 1963, **67**, 527–534.

Clark, K. B. *Dark ghetto*. New York: Harper & Row, 1965.

Coleman, J., **Campbell**, E., **Hobson**, C., **McPartland**, I., **Mood**, A., **Weinfeld**, F., & **York**, R. *Equality of educational opportunity*. Washington, D.C.: United States Government Printing Office, 1966.

Crandall, V. C., **Crandall**, V. J., & **Katkovsky**, E. A childrens' Social Desirability Questionnaire. *Journal of Consulting Psychology*, 1965, **29**, 27–36.

Epstein, R. Need for approval and the conditioning of verbal hostility in asthmatic children. *Journal of Abnormal and Social Psychology*, 1964, **69**, 105–109.

Gottlieb, D. Teaching and students: The views of Negro and white teachers. In S. Webster (Ed.), *The disadvantaged learner*. San Francisco: Chandler, 1964.

Katz, I. Review of evidence relating to effects of desegregation on the intellectual performance of Negroes. *American Psychologist*, 1964, **19**, 381–399.

Katz, I., **Henchy**, T., & **Allen**, H. Effects of race of tester, approval-disapproval, and need for approval on Negro childrens' learning. *Journal of Personality and Social Psychology*, 1968, **8**, 38–42.

Rosenhan, D. Effects of social class and race upon responsiveness to approval and disapproval. *Journal of Personality and Social Psychology*, 1966, **4**, 253–259.

Marlowe, D., & **Crowne**, D. New scale of Social Desirability. *Journal of Consulting Psychology*, 1960, **24**, 349–354.

Stevenson, H. W. Social reinforcement of childrens' behavior. In L. P. Lipsitt & C. Spiker (Eds.), *Advances in Child Development and Behavior: II*. New York: Academic Press, 1965.

Warner, L., **Meeker**, M., & **Eells**, K. *Social class in America: A manual for procedure for the measurement of social status*. Chicago, Ill.: Science Research Associates, 1949.

(Received March 13, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 3, 245–252

REINFORCEMENT AND EDUCATIONALLY DISADVANTAGED BOYS:

AN EXPLORATORY STUDY[1]†

ANTHONY F. FAZIO,[2] JEFFREY M. LOHR,[3] AND C. G. SCREVEN
University of Wisconson, Milwaukee

The present investigation evaluated "intrinsic motivation" of educationally disadvantaged boys to make visual reading discriminations. The task was presented in a noisy recreation room of an inner-city Boys' Club. Extrinsically imposed reinforcement contingencies were not employed and subjects were free to not participate. The findings were as follows: 311 boys participated between 1 and 21 times (averaging 24 responses per game); average number of responses did not diminish with repeated participations; a median of 40% of those in the room during each session participated; and, more than 1,980 games were played. The findings were related to typical behavior modification orientations used in special education programs.

The application of operant conditioning principles to the development of techniques to motivate and instruct the resistant learner, the underachiever, and the retardate have produced some very encouraging results (e.g., Staats, Finley, Minke, & Wolf, 1964; Staats, Minke, & Butts, 1971; Staats, Minke, Goodwin, & Landeen, 1967; Wolf, Giles, & Hall, 1968). This research, in general, may be characterized as employing relatively highly structured, institutionalized settings (be they classroom or laboratory) and "extrinsic" reinforcements to maintain attendance and performance. Often consumable and almost always tangible reinforcements are used (e.g., candy, tokens, trinkets, slips of paper). In part this may be the result of approaches based upon research using infrahumans. In part it may be also the result of an educational misconception, as exemplified by Ausubel (1966) who suggested that "... culturally deprived children typically manifest little intrinsic motivation to learn [p. 473]." Experimental support for this scientific prejudice consists of reports (e.g., Terrell, Durkin, & Wiesley, 1959) that lower-class children learned faster when given tangible reinforcers such as candy or plastic chips, and middle-class children learned faster when given "abstract" reinforcers such as a flash of light. Rosenhan and Greenwald (1965) and Safer and Kornreich (1968), however, failed to replicate these results. Thus the use of tangible reinforcers to teach lower-class children may not be necessary.

A similar misconception may exist with respect to techniques for motivating lower-class children. Token reinforcement systems have been designed (e.g., Staats, 1968) to maintain voluntary participation in complex reading discriminations over long periods of time. However, tangible reinforcers may not be necessary to obtain and maintain such behaviors. Premack (1959, 1961) and Allyon and Azrin (1968) have suggested that the probability of occurrence of a given response serves as an index of its reinforcing value. The key to uncovering long-lasting, effective reinforcers is to observe what the subjects do when they are free to do several things but not required to do any one thing; that is, when they are not under an extrinsically imposed schedule of reinforcement. It would be these behaviors that come closest to being classified "natural" reinforcers.

The purpose of the present project was

---

[1] This research was supported in part by a grant from the graduate school of the University of Wisconsin at Milwaukee. The authors wish to thank Wilbur Davis and Roy Hill, Directors of the LaVarnway Unit of the Milwaukee Boys' Club for their suggestions and cooperation; Edward Giese for assembling the apparatus; Michael Flaherty, Daniel Hendricks, Fred Jefferson, and Wesley Rouse for assisting with the project; and, Arthur W. Staats for his comments on the manuscript.

[2] Requests for reprints should be sent to Anthony F. Fazio, Department of Psychology, University of Wisconsin, Milwaukee, Wisconsin 53201.

[3] Now at the University of Hawaii.

first to see if educationally disadvantaged youths would engage in educational tasks without coercion or extrinsic reinforcement, and, second to explore the motivational possibilities of an automated instructional system. The system was designed to function as a highly motivating activity, be automated and free standing, employ "natural" reinforcers (such as might be involved in winning a game of Ping-Pong), and elicit difficult visual-reading responses.

METHOD

Subjects

During the 9 months the study was in progress a median of 34 7- to 13-year-old boys of a central city Milwaukee Boys' Club with predominantly black membership attended a recreation room on a daily basis. The actual number of participants in the learning task game was 311. There were approximately 9,565 members of the Club in this age group during any 1 month. Approximately one out of four boys from the area came from a broken home, and a large number of families were supported through public welfare (Milwaukee Services Development Committee, 1966).

Apparatus

The present study involved four experimental conditions.[4] In the first two conditions, the apparatus consisted of two parallel pretests and posttests and a visual matching-to-sample program (referred to as the word game). Each pretest and posttest consisted of 12 pages from the Smith (1967) Michigan Language Program. The subject was given the pretest: 12 paper sheets, each containing a paragraph of sentences or phrases. At the top of the paragraph was a sample phrase which the subject was to find within the paragraph below and circle it. The actual Game consisted of 203 sheets, each page containing a four-choice matching-to-sample problem modified from Levels 6 through 8 of the Michigan Language Program. The sample was written correctly, and the four possible answer choices consisted of the sample word spelled correctly and three similar but misspelled words (e.g., boat, doat, doet, boad). The 203 reading frames were divided into four approximately equal sections or "quarters." In the first condition a black male assistant administered the program; in the second, a white male administered the program.

The third experimental condition used a parallel form of the program used in Conditions 1 and 2. The problems were transcribed onto fanfold program paper and inserted in an MTA S-R-400 stimulus presentation device equipped with a four-choice press-panel cover which allowed each word to appear directly behind a response key. The press-panel response mode was chosen because of the desirability of stimulus cue, response, and reinforcement continguity in complex visual discriminations (cf. Cowey & Weiskrantz, 1968).

The fourth experimental condition made use of this same MTA S-R-400 system, but it was set vertically within a large chassis somewhat resembling a penny arcade device. The top-half of the front of the device was a large white Plexiglas panel behind which were mounted two large pinball machine type counters and two columns of red lights. Below this panel rested the S-R-400. To either side of the S-R-400 were two smaller panels, one with a red light behind it and one with a blue light behind it. Players were then designated either "red" or "blue," each with his own counter and column of lights.

Procedure

Important aspects of the procedure were the following: the subjects were free to participate in any of the many activities and games available in the recreation room; no deprivation procedures were instituted to insure the motivation of the boys; reading games were seldom used in the recreation room before the experiment began; the task was not a prerequisite for any other activity in or out of the room; and, the assistants were instructed to respond only neutrally whether or not a boy expressed interest in or participated in the task.

The four conditions were run sequentially (cf. Antonitis & Barnes, 1961; Antonitis, Frey, & Baron, 1964). The experiment took place in a recreation room of the Boys' Club that was 78 feet long × 29 feet wide. In the same room were pool tables, Ping-Pong tables, assorted boxed games, and a color television. In adjacent rooms other activities (e.g., basketball, swimming) were available. In all four conditions, the subjects signed their names and Boys' Club card number on a sign up sheet before being pretested or playing the game. Experimental sessions ran from 5:00 P.M. to 7:00 P.M., Monday through Friday except for holidays, apparatus failure, and room maintenance. The entire experiment ran for 9 months (September through June).

Conditions 1 and 2. After signing up, the subject's record was checked to see if he previously had been pretested. If so, the record was checked to see what quarter of the game he had played previously. Subjects were instructed to find and touch one of the four choices which matched the sample at the top of each page. They were told they could play the game as long as they wished. Those playing for the first time began with the first quarter of 50 frames, subsequently playing through Quarters 2, 3, and 4 if they chose. If a subject completed a quarter in a previous game he could start the next game at or below that level, providing his "opponent" could so qualify.[5]

---

[4] A set of 35 mm. color transparencies of the experimental setting and apparatus is on file in the library of the University of Wisconsin at Milwaukee.

A coin was tossed to determine which subject would respond first. Following either a correct or incorrect response, it became the opponent's turn to respond. After correct responses, the assistant presented a new frame (i.e., flipped the page). After incorrect responses, a new frame was not presented until one of the subjects made a correct response. Feedback and "reinforcement" were provided by the assistant who said only "correct" or "incorrect" after each response. The assistant also tallied the number of correct and incorrect responses for each subject. At the end of the fourth quarter, the subject with the most correct responses was declared the "winner." If one of the subjects quit before the completion of the fourth quarter, no winner was announced and the remaining subject was allowed to start another game with a new opponent. The level at which the new game began was determined by the mutual consent and qualifications of the opponents. The subject played until one or both wished to leave.

After 26 sessions, the first assistant terminated employment. In order to investigate possible racial effects of the assistant, a Caucasian was hired. Condition 2 began with this new assistant using the apparatus (i.e., booklet of frames) described for Condition 1. The second assistant administered the second, third, and fourth conditions. With the second assistant, the procedure of limiting each subject to playing four quarters in one sitting was initiated. A subject could play more than four quarters in one sitting only if there were no other boys who wished to play the game. Condition 2 was terminated after 29 sessions.

Condition 3. Condition 3 was initiated to investigate possible motivational aspects of an automatic device which the subjects could manipulate (cf. Bijou & Sturges, 1959; Screven, 1969). A parallel-form 200-frame game was transcribed onto fanfold paper and programmed to advance to the next frame only on a correct response. The S-R-400 was placed in the same location as the apparatus in Conditions 1 and 2. The pretest procedure continued for seven sessions as before and then it was terminated. The number of subjects signing to play increased compared with Conditions 1 and 2 and the pretest procedure limited the number of boys who could play in the 2-hour session. Procedures carried out by the assistant were the same as before except for those functions assumed by the S-R-400. "Turns" were alternated after each response, right and wrong answers were tallied by the assistant, and after four quarters, the winner was announced. This time, however, feedback and sensory reinforcement were provided by the automatic changing of the frames and not by the assistant. After 35 sessions, Condition 3 was terminated.

Condition 4. Condition 4 was introduced to investigate the possible motivational aspects of a device resembling an amusement or penny arcade machine. As in Condition 3, answer choice was indicated by pressing the appropriate choice panel, and only a correct choice advanced the frames. Other contingencies were different, however. Each subject had a different colored panel on his side of the apparatus which when illuminated was a discriminative stimulus for a response. Three seconds after any response the subject's panel darkened and his opponent's panel was illuminated, cueing alternation of responses. A subject's score was tallied by the large "pin-ball" counter on either side of the faceplate. A correct response advanced the counter one point. An incorrect response had no effect on counter points, although it did give the opponent a greater probability of making a correct response and of gaining a point. The subjects could not affect the number of points on their opponent's counter. Smaller counters not visible to the subjects tallied the total number of responses for each subject. Condition 4 was terminated after 28 sessions when the Boys' Club summer schedule went into effect.

Dependent Measures

For each session (i.e., day) signatures on the sign-up sheet, number of players, and total number of responses per player provided measures of interest in the game. During each session the number of boys in the room was counted every ½ hour. The ratios of each of these measures to mean head count per day were plotted in blocks of three sessions for experimental conditions. These ratios controlled for daily fluctuations in recreation room attendance due to the weather, vacations, etc. Records were kept for each boy participating in the game and contained the number of correct, incorrect, and total responses for each time the game was played. Pretest and posttest data were not examined in the present study.

RESULTS

Because data were collected from boys who participated only once (i.e., independent groups) as well as from boys who participated as many as 21 times within and across experimental conditions (i.e., repeated measures) no parametric statistical analyses of the data appeared to be appropriate. The interpretation of operant behaviors of groups of subjects has sometimes been a problem for researchers, and usually only descriptive material is presented (cf. Antonitis & Barnes, 1961; Antonitis, Frey, & Baron, 1964). The significance and reliability of the data obtained here were inferred from sign tests whenever possible. In

---

[5] In many cases the subjects acted as partners and gave their "opponents" the correct answers.

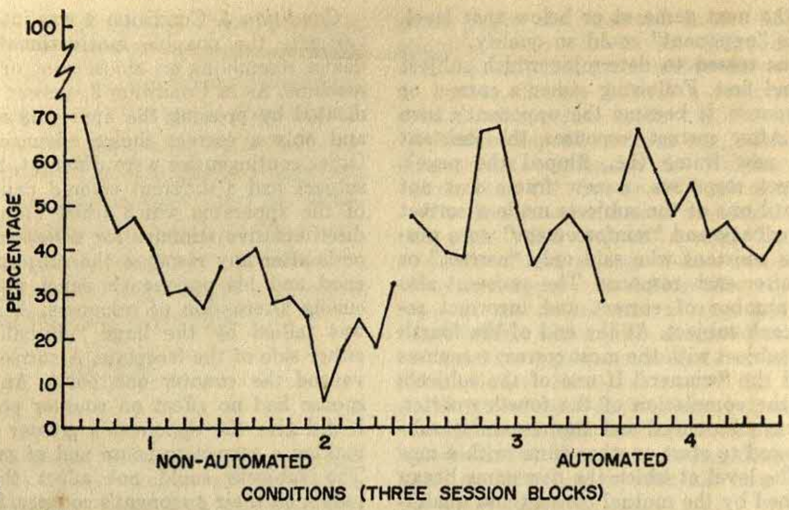


Fig. 1. Percentage of boys in the recreation room signing to participate. (The number of boys in the room per session averaged 25.3 for Condition 1, range 21.5–28.6; 32.9 for Condition 2, range 26.6–45.1; 38.0 for Condition 3, range 31.9 to 42.3; and 31.1 for Condition 4, range 26.4–39.3.)

addition, the number of observations represented by data points (e.g., Figure 2) and the consistency of the differences between experimental conditions (e.g., Conditions 1 and 2 versus 3 and 4 in Figures 1, 2, and 3) appeared to strengthen inferences made from the data. The ratio of boys in the recreation room who signed to participate during each session were graphed as percentages. Inspection of Figure 1 revealed that during several sessions more than half of the boys in the room expressed an interest in participating. On many occasions boys were lined up to participate and expressed concern about losing their place, others cutting in, etc.

Each experimental condition was divided into first and second halves (for odd numbers of blocks the middle session was ignored). A total of 20 first-half scores were matched and compared with second-half scores using a sign test. Interest was found to have dwindled significantly ($p = .012$) within apparatus conditions (Figure 1).[6]

Comparing the first two points of each condition with the last two points of the previous condition revealed that interest was briefly stimulated by the introduction of a new experimental condition and/or apparatus ($p = .032$; the Hawthorne effect; Figure 1).

Again using a sign test, the automated conditions (i.e., 3 and 4) were compared with the nonautomated conditions (i.e., 1 and 2). (Two data points from Condition 3 and one from Condition 4 were randomly discarded to make 19 matched pairs.) The automated conditions were found to have sustained a higher level of interest than did the nonautomated conditions ($p = .02$; Figure 1).

Average number of total responses per player per day were plotted in Figure 2. Although this measure of interest correlated significantly ($r_{pm} = .53, p < .001$) with the sign-up data, and differences within and between experimental conditions were identical to those reported for the sign-up data, only the automated–nonautomated comparison achieved statistical significance ($p < .004$).

Average numbers of players per day are presented in Figure 3. This measure was also found to correlate significantly with both the sign-up data ($r_{pm} = .78, p < .001$) and the average number of total response data ($r_{pm} = .53, p < .001$). Again the differences obtained were similar to those described above. Interest was again found to have dwindled significantly ($p = .02$)

---

[6] **All p values are for two-tailed tests.**

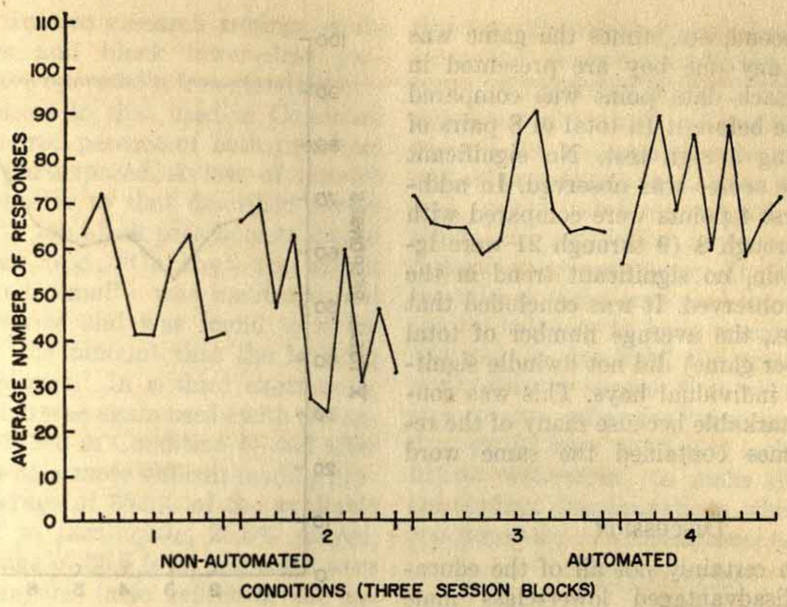


FIG. 2. Average numbers of total responses per player per day. (Each data point represents different numbers of games played; the median number of observations and range represented by each point for each condition were as follows: For Condition 1 median = 10, range 6–15; for Condition 2 median = 14.5, range 3–16; for Condition 3 median = 25.5, range 12–38; and for Condition 4 median = 21, range 5–28.)

within conditions, and the automated conditions were found to have attracted a higher percentage of players than the nonautomated conditions ($p < .001$). The Hawthorne effect, although similar to that described for the other measure was not statistically significant.

Individual differences in participating were of course considerable. Figure 4 reveals the distribution of number of games played. With respect to these participations the mode was 1; median, 2.2; and mean, 2.3.

The mean numbers of total responses for

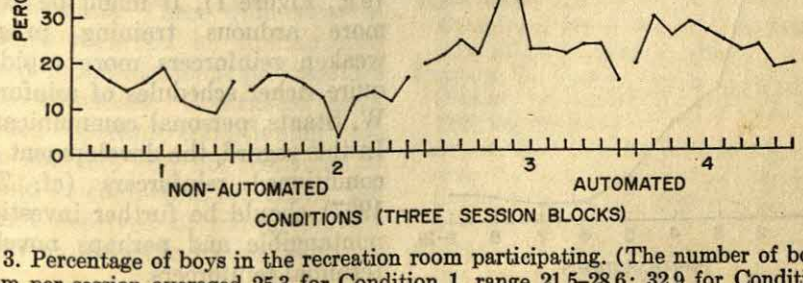


FIG. 3. Percentage of boys in the recreation room participating. (The number of boys in the room per session averaged 25.3 for Condition 1, range 21.5–28.6; 32.9 for Condition 2, range 26.6–45.1; 38.0 for Condition 3, range 31.9–42.3; and 31.1 for Condition 4, range 26.4–38.3.)

the first, second, etc., times the game was played by any one boy are presented in Figure 5. Each data point was compared with the one before it (a total of 8 pairs of points) using a sign test. No significant trend in the scores was observed. In addition, the first 4 points were compared with points 5 through 8 (9 through 21 were ignored). Again, no significant trend in the scores was observed. It was concluded that interest (i.e., the average number of total responses per game) did not dwindle significantly for individual boys. This was considered remarkable because many of the repeated games contained the same word stimuli.

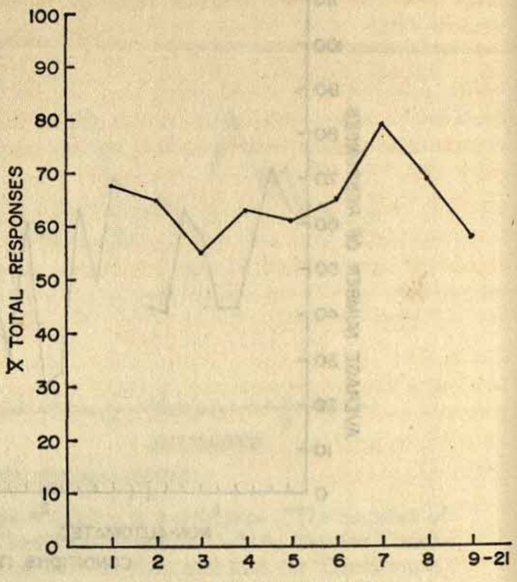


Fig. 5. Average numbers of total responses per participation (*n* decreasing; see Figure 4).

Discussion

Although certainly not all of the educationally disadvantaged lower-class male members available were motivated to respond in a difficult visual discrimination task involving spelling, a median of 40% of those present in the recreation room made an average of 24 such responses over a total of 1,980 games during 118 sessions. Three hundred and eleven boys participated. There were, it should be emphasized, new billiard tables, a colored television set, etc., available as competing alternatives. The significance of this interest and participation is further enhanced when it is realized that "spelling words" appeared to command very little masculine peer approval compared to playing pool or basketball. The only reinforcement used for performance involved physical and social stimuli intrinsic to participating in any recreational activity.

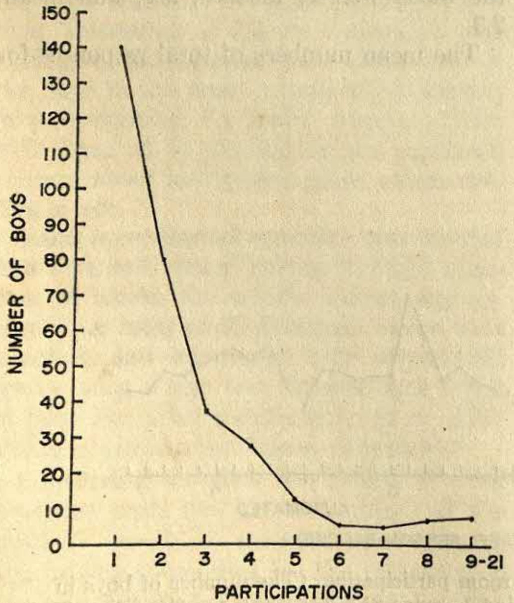


Fig. 4. Frequency distribution for number of total participations by individual boys.

Other types of reinforcers (e.g., points in a tournament; tokens exchangeable for summer camp facilities) may attract larger numbers of boys and maintain interest for longer periods. This remains to be investigated. In this regard, "weakening" of the reinforcers was apparent within conditions (e.g., Figure 1). It might be expected that more arduous training programs will weaken reinforcers more rapidly and require richer schedules of reinforcement (A. W. Staats, personal communication, 1970). In this regard, the development of powerful conditioned reinforcers (cf. Zimmerman, 1957) should be further investigated using nontangible and perhaps novel events as terminal reinforcers.

In three related experiments, still in progress, these findings have been replicated in large part. In two research settings white middle-class and black lower-class preschoolers have operated a free-standing machine (identical to that used in Condition 3). One hundred percent of both preschool classrooms participated. A loss of interest over time similar to that described earlier occurred with the black preschoolers. Social reinforcement (e.g., "Oh boy! You got it right the first time!") was instituted and withdrawn twice and was found to affect ($p = .016$) the amount that the learning game was played.[7] In a third experiment, the Boys' Club was again used (with the apparatus described in Condition 4) and after 15 sessions with a more difficult reading program an average of 36.5% of the available boys signed to participate, 25.6% played, and an average of 38.6 total responses were made by players (also reflecting the increased difficulty of the program).

One implication drawn from these results concerns the potential relative inexpensiveness of making instructional systems available to students during lesiure and/or recreation periods. The devices can be free standing and automated. A second area for further investigation concerns opportunities for mutual teaching and reinforcement, or "social facilitation." Rather than promoting competitive attitudes, educational cooperation might involve the valuable service of having more advanced students practice the material while helping less advanced students. In the present experiment, cooperation occurred spontaneously and could easily be reinforced to further encourage "helping" attitudes.

In most aspects the present experimental procedure was unlike typical behavior modification approaches based on research involving infrahumans. These differences in procedure were regarded as fundamental to educational–rehabilitational philosophies and deserve emphases. Contrasted with more orthodox operant conditioning approaches, the subjects were not restricted to an area devoid of competing activities; the boys were not deprived of reinforcers involved in playing games to insure their motivation; the reading responses in question more than likely did not have a high probability of occurring inside or outside of the experimental setting; no one was required to emit the response in order to engage in any other activity (e.g., leave the setting); and, social reinforcement, if any, from adults did not follow responses with any immediacy or consistency. In short, most of the rules or techniques used to establish and control behavior in basic research were not followed. In this light, the present findings were taken to suggest that when given the opportunity, otherwise "resistant learners" as a group were not found lacking in "intrinsic motivation" to make difficult visual-reading discriminations when the task was presented as a recreational activity.

---

[7] A detailed report of this experiment is in preparation.

REFERENCES

ALLYON, T., & AZRIN, N. *The token economy: A motivational system for therapy and rehabilitation.* New York: Appleton-Century-Crofts, 1968.

ANTONITIS, J. J., & BARNES, G. W. Group operant behavior: An extension of individual research methodology to a real-life situation. *Journal of Genetic Psychology,* 1961, **98,** 95–111.

ANTONITIS, J. J., FREY, R. B., & BARON, A. Group operant behavior: Effects of tape recorded verbal reinforcers on bar-pressing behavior of preschool children in a real life situation. *Journal of Genetic Psychology,* 1964, **105,** 311–331.

AUSUBEL, D. P. A teaching strategy for culturally deprived pupils: Cognitive and motivational considerations. In S. W. Webster (Ed.), *Educating the disadvantaged learner.* San Francisco: Chandler, 1966.

BIJOU, S. W., & STURGES, P. T. Positive reinforcers for experimental studies with children—consumables and manipulatables. *Child Development,* 1959, **30,** 151–170.

COWEY, B. M., & WEISKRANTZ, L. Varying spatial separation of cues, response and reward in visual discrimination in monkeys. *Journal of Comparative and Physiological Psychology,* 1968, **66,** 220–224.

MILWAUKEE SERVICES DEVELOPMENT COMMITTEE. *Guidelines for planning.* Milwaukee, Wisc.: Boys' Clubs of America, 1966.

PREMACK, D. Toward empirical behavior laws: I. Positive reinforcement. *Psychological Review,* 1959, **66,** 219–223.

PREMACK, D. Predicting instrumental performance from the independent rate of the contingent response. *Journal of Experimental Psychology,* 1961, **61,** 163–171.

ROSENHAN, D., & GREENWALD, J. A. The effects of age, sex, and socioeconomic class on responsiveness to two classes of verbal reinforcement. *Journal of Personality,* 1965, **33,** 108–121.

Safer, M. A., & Kornreich, L. B. The interaction of social class and type of reinforcement in discrimination learning. *Psychonomic Science,* 1968, **11,** 206.

Screven, C. G. The museum as a responsive learning environment. *Museum News,* 1969, **48,** 7–10.

Smith, D. E. P. (Ed.). *Michigan language program.* Ann Arbor, Mich.: Ann Arbor, 1967.

Staats, A. W. A general apparatus for the investigation of complex learning in children. *Behaviour Research and Therapy,* 1968, **6,** 45–50.

Staats, A. W., Finley, J. A., Minke, K. A., & Wolf, M. M. Reinforcement variables in the control of unit reading responses. *Journal of the Experimental Analysis of Behavior,* 1964, **7,** 139–149.

Staats, A. W., Minke, K. A., & Butts, P. A token reinforcement remedial reading program administered by black instructional-technicians to backward black children. *Behavior Therapy,* 1971, in press.

Staats, A. W., Minke, K. A., Goodwin, W., & Landeen, J. Cognitive behavior modification: "Motivated learning" reading treatment with subprofessional therapy-technicians. *Behaviour Research and Therapy,* 1967, **5,** 283–299.

Terrell, G., Durkin, K., & Wiesley, M. Social class and the nature of incentive in discrimination learning. *Journal of Abnormal and Social Psychology,* 1959, **59,** 270–272.

Wolf, M. M., Giles, D. K., & Hall, R. V. Experiments with token reinforcement in a remedial classroom. *Behaviour Research and Therapy,* 1968, **6,** 51–64.

Zimmermann, D. W. Durable secondary reinforcement: Method and theory. *Psychological Review,* 1957, **64,** 373–383.

(Received March 14, 1970)

JOURNAL OF
EDUCATIONAL PSYCHOLOGY
MONOGRAPH

Vol. 62, No. 3 June, 1971

TEACHING CONCEPTS IN THE CLASSROOM: A SET OF TEACHING PRESCRIPTIONS DERIVED FROM EXPERIMENTAL RESEARCH

D. CECIL CLARK[1]
University of Washington, Seattle

This paper attempts to answer the following question: To what extent can the large body of experimental research in concept attainment be used to suggest an empirically derived set of prescriptions for teaching certain types of classroom concepts? From the majority of experimental research on concept attainment over the past 35 years, 235 studies were selected whose independent variables could be identified and placed into the following groups: concept variable, stimulus variables, subject variables, and task variables. Section I delineates factors in the experimental context which may lower external validity. Section II provides four examples of conceptual objectives found in the classroom, along with their operational evidence. Section III includes a set of 61 statements about concept attainment formulated from the research, while Section IV is a set of teaching prescriptions suggested by these statements. Section V cites the research from which each statement was derived, as well as research which failed to support that statement.

Over the past 30 years there have been well over 250 experimental studies conducted in the area of concept attainment. Examples of independent variables in these experiments are the type of concept, stimulus dimensions, subject characteristics, and task sequences. The vast majority of these studies are highly specific and carefully controlled, with internal validity receiving primary attention. A typical experiment involves a subject being presented with sets of geometric patterns which vary along several dimensions. His task is to sort, usually with feedback, until he discovers the "correct" critical dimensions. He then correctly sorts on to some predetermined criterion, thereby demonstrating attainment of the concept.

Paralleling this vigorous experimental activity, one observes the great number of concept-formation tasks filling a major part of most public school curricula. Daily, teachers at every age level present a variety of concepts to from 20 to 35 children. The materials and procedures used by several teachers, each presenting the same concept to his own class, represent startling variation, as does the same teacher presenting two different concepts to one class. Additionally, teachers appear to have considerable difficulty mustering rationale for the materials they select and procedures they follow in any given concept-presenting experience. With this amount of variation and dearth of rationale, some if not many of the procedures followed in classroom concept formation must be less than effective.

In this article, an attempt is made to answer the following question: To what extent can the large body of experimental research in concept attainment be used to suggest an empirically derived set of prescriptions for teaching certain types of classroom concepts?

The article is divided into five sections. The first delineates factors which might possibly lower external validity (generalizability) of research results when applied to the classroom setting. The strength of subsequent sections depends upon the seriousness of these factors. The second section contains four examples of conceptual objectives that might be selected by the teacher and achieved more optimally by the student, presumably through systematic application of the research. Section III details a set of statements about concept attainment that have been derived from the research. The statements are classified according to the following categories of independent variables: concept variable, stimulus variables, subject variables, and task variables.[2] Section IV is a set of teaching prescriptions suggested by the statements, and it includes an example showing how these prescriptions can be developed for a given concept. Finally, Section V cites the research from which each statement was derived, as well as research which failed to support that statement.

---

[1] Requests for reprints should be sent to D. Cecil Clark, College of Education, University of Washington, Seattle, Washington 98105.

SECTION I. FACTORS THAT MAY AFFECT EXTERNAL VALIDITY

There are at least two factors to be considered. The first is the degree to which the two major parameters of every concept task can be identified in the classroom experiences. Since combinations of these two parameters have served as independent variables in the great majority of the research, and since their influence on concept attainment is largely known, they must be continually and completely identified in the classroom concept experiences in order to prescribe modifications. These two parameters are concept instances and dimensions displayed by the instances. Instances are either positive or negative, and dimensions either critical or noncritical.

The second factor is the degree to which an independent variable affecting concept attainment in the context of an experimental study maintains its effect when transplanted to the context of the classroom. There are at least five areas in which the two contexts differ relative to concept experiences, and these differences may reduce or enhance the effect of a given independent variable as it is transplanted to the classroom context. These areas are as follows:

1. Type of concept
   A. Research context
      1. The substantial majority of studies have utilized conjunctive concepts;[3] fewer studies have dealt with unidimensional (primitive), disjunctive, relational, and biconditional concepts.
   B. Classroom context
      1. At the elementary and secondary levels, the great majority of concepts appear to be conjunctive and relational in nature.

II. Goal of the concept task
   A. Research context
      1. The subject is required to attain an already familiar concept;[4] familiar implicit responses are evoked.
      2. The subject is frequently required to attain several concepts simultaneously.
   B. Classroom context
      1. The student is required to form a new and unfamiliar concept; nonfamiliar implicit responses are evoked.
      2. The student typically forms one concept at a time in a successive fashion.

III. Instances and dimensions of the concept
   A. Research context
      1. Instances are almost exclusively concrete (physical objects).

---

[2] An attempt was made to limit the review of research to those studies—mostly experiments—in which the independent variables could be generally identified and were in some way varied. This delineation tends to not only create stronger internal validity (control) for establishing the statements, but also permits them to be classified in terms of independent variables, with concept "attainment" as the dependent variable. Experiments in some areas were omitted because their independent variables were inappropriate or overly confusing when translating into the classroom context. Examples are searching strategies used by the subject in a discovery paradigm, reversal and nonreversal shifting, and relevant and irrelevant redundancy.

[3] More technically, they have used concepts whose critical dimensions are related according to a conjunctive rule.

[4] More technically, he is required to discover the conceptual rule relating already familiar dimensions.

  2. Instances display dimensions (e.g., color) each of which has two or more values (e.g., black–white).
  3. The values along a dimension are completely discrete in nature (e.g., black or white).
  4. The number of dimensions is finite as are their possible combinations (e.g., if color, black–white and if size, large–small are the only two observable dimensions, then there are but four possible combinations, black–large, black–small, white–large, white–small).
  5. The critical dimensions are absolute; that is, if the color *white* is critical, its whiteness and criticality remain the same over time (unless, of course, the experimenter intentionally changes the criticality of a dimension).
- B. Classroom context
  1. Instances become increasingly abstract (verbal) as the student moves up through the various grade levels.
  2. Instances display properties (e.g., a leaf). Each property can be viewed along several dimensions (e.g., its color, size, texture, etc.). Each dimension can have any number of values (e.g., light green to dark green).
  3. The values along a dimension are virtually continuous in nature (e.g., light green to dark green).
  4. The number of possible dimensions is infinite as are the possible combinations (e.g., see leaf above).
  5. The critical dimensions are relative and may change over time; that is, a critical dimension in the early grades usually requires qualification and greater sophistication in the later grades. Thus, because the critical dimensions change, the concept changes.

IV. Concept task
- A. Research context
  1. A discovery strategy is used in which the critical dimensions are to be discovered by the subject.
- B. Classroom context
  1. A reception strategy is most frequently used in which the teacher points out the critical dimensions to be learned.

V. Evaluation of concept attainment
- A. Research context
  1. Concept attainment is most often measured by a simple sorting task telling the experimenter whether the concept has or has not been attained.
- B. Classroom context
  1. Concept formation, when measured, is usually done by asking the student to "define" or "use" the concept. At its best, the measurement procedure tells the teacher the degree to which the concept has been formed.

Whether or not these contextual differences significantly confound transplanted independent variables can only be determined through a series of empirical investigations in classroom-type settings. One such series is being currently conducted by the author.

SECTION II. EXAMPLES OF CONCEPTUAL OBJECTIVES

In the vast majority of the studies surveyed, concept "attainment" is signified by the subject's ability to perform some type of sorting task. In its simple form, for example, he is presented with a variety of geometric figures and asked to sort all the "squares" into one pile and the "nonsquares" into another. When he is able to perform this task correctly over a series of trials, he has "attained" the concept.

As one examines the dependent variable, concept attainment, in the classroom setting, at least two differences arise. First, a wider variety of less precisely defined behavior is accepted as evidence of concept attainment. Second, when measurement of concept attainment does occur, verbal definitions of the concept rather than sorting tasks often serve as primary evidence.

In order to maximize the possible application of the research to the classroom—at this

TABLE 1

EXAMPLES OF CONCEPTUAL OBJECTIVES AND THEIR OVERT EVIDENCE

1 Objectives stated in covert, familiar educational form	2 Objectives stated in covert, technical form	3 Overt behavioral evidence signifying attainment of the objective
The student will have:	The student will have:	The student will be able to:
1. Memorized a definition of the concept.	1. Memorized the critical properties in appropriate relationship.[a]	1. Verbally recall the critical properties in appropriate rerelationship.
2. Formed the concept through direct instruction.	2. Abstracted and generalized the critical properties in appropriate relationship through a reception strategy.	2. Correctly sort out a variety of positive and negative instances of the instructed concept.
3. Applied the concept to a new situation.	3. Identified new and unfamiliar positive and negative instances of the concept as a result of examining their critical properties in relationship.	3. Given any new and unfamiliar positive or negative instance, verbally identify critical and noncritical properties and the appropriateness or inappropriateness of their relationship.
4. Discovered the concept in the absence of instruction.	4. Abstracted and generalized unfamiliar critical properties in appropriate relationship through a discovery strategy.	4. Correctly sort out a variety of positive and negative instances of the discovered concept.

[a] More technically, the student will have memorized the critical properties according to their combinatorial rule.

point in terms of the similarity of dependent variables in both settings—concept attainment in the classroom needs to be more sharply and operationally defined, and it should probably include some form of sorting task as at least one operational measure. One procedure for accomplishing this was as follows:

The *Taxonomy of Educational Objectives: Cognitive Domain* (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) was examined. The taxonomy represents an exhaustive list of cognitive operations the student is likely to engage in at some point in his formal education. These covert behaviors have been placed in an assumed hierarchy of complexity. For the present paper, they were used with concepts only. Thus, the student *knows* concepts, *comprehends* concepts, *applies* concepts, *analyzes* concepts, *synthesizes* concepts, and *evaluates* concepts.

Based on an analysis of curriculum guides, teacher's editions of textbooks, and anecdotal classroom observations, the frequency with which each of these behaviors, relative to concepts, tended to occur was determined. *Knowing*, *comprehending*, and *applying* concepts appeared with the greatest frequency. Also, "discovering new concepts on his own"—which is most closely related to *synthesis*—appeared with some frequency. These four covert behaviors, then, were stated as objectives in terms commonly used by teachers (Table 1, column 1). For example, a teacher often declares, "My students should be able to apply this concept to new situations." Next, these objectives were translated, still in covert form, into the more technical language used in the research studies (column 2). Finally, they were operationalized as overt behavioral tasks (column 3). Following is an example of how the four objectives (Table 1) might be stated in terms of a particular concept, "square."

Objective stated in covert, familiar educational form (column 1). The student will have (1) memorized the definition of a "square"; (2) formed the concept "square" through direct instruction; (3) applied the concept "square" to a new situation; (4) discovered the concept "square" in the absence of instruction.

Objective stated in covert, technical form (column 2). The student will have (1) memorized the critical properties of a "square" in appropriate relationship; (2) abstracted and generalized the critical properties of a "square" in their appropriate relationship through a reception strategy; (3) identified new and unfamiliar positive and negative instances of a "square" as a result of examining their critical properties in relationship; (4) abstracted and generalized the critical properties of a "square" in their appropriate relationship through a discovery strategy.

Overt behavioral evidence signifying attainment of the objective (column 3). The student will be able to (1) verbally recall critical properties of a "square" in their appropriate relationship; (2) given a group of geometric figures, correctly sort out "squares" from "nonsquares"; (3) given a new and unfamiliar "square" or "nonsquare," verbally identify its critical and noncritical properties and the appropriateness or inappropriateness of their relationship; (4) given a group of geometric figures, correctly sort out "squares" from "nonsquares."

Procedurally, a teacher might select any one or more of the objectives in column 1 in preparation for teaching about "squares." She then moves over to column 3 with each selection to determine its overt behavioral evidence.

The four conceptual objectives in Table 1 simply represent examples of those that are frequently encountered in the typical classroom setting; they are not meant to be exhaustive. There is little empirical evidence justifying their assignment into any clear hierarchy of complexity. Furthermore, since most of the surveyed research bears heavily on the type exemplified by Number 4, caution must be exercised in generalizing results to the other three types. Finally, these particular overt behavioral tasks (column 3) may not be different enough to justify the listing of distinctly different objectives (column 1). However, the presentation of a well-developed taxonomy of cognitive objectives is not the intent of this paper, and the above limitations need not interfere in generating a set of teaching prescriptions as long as at least one operational definition of concept attainment for the classroom involves a sorting task.

SECTION III. STATEMENTS ABOUT CONCEPT ATTAINMENT

As a first step, all of the surveyed studies with similar independent variables were grouped together. For example, independent variables in three separate studies were (1) conjunctive versus disjunctive concepts, (2) conjunctive versus relational versus biconditional concepts, (3) concepts of concrete objects versus concepts of spatial forms versus concepts of numbers. These studies were placed into the same category which emerged with the label *type of concept* as the independent variable. Other categories that emerged were stimulus variables, subject variables, and task variables. Subcategories then began to emerge within each general category. Thus, in the above example, Studies 1 and 2 were ultimately placed in one subcategory while Study 3, along with others similar in nature, were placed in a second subcategory. As might be expected, the number of studies in a category varied. Where there was a large number of studies in a category, more subcategories emerged and, in many cases, these subcategories themselves were further subdivided. Attempts were continually made to place only those studies within a subcategory which were as homogeneous as possible in terms of their independent variables. While this effort acted to purify a subcategory, it also tended to reduce the number of studies per subcategory.

Next, results within each subcategory were examined and the most frequently occurring result was formulated as a statement about concept attainment. The number of results failing to support the statement was also recorded (see Statements below). For example, out of 41 studies falling within one of the subcategories, the results of 40 suggested that, as the number of noncritical properties decreases, ease of concept attainment increases; only one study suggested nonsupportive results (Statement 11). A *Statement about Concept Attainment* is here defined as a formalization of the majority results of studies within a subcategory. Those subcategories with results which were overly confusing or equally contradictory were omitted. On the other hand, some categories were retained with the results of but one study because of their greater potential relevance to the educational setting.

Understandably, a statement emerging from the results of only one study is more tentative and, at present, receives less attention than a statement emerging from as many as 40 studies. Likewise, a statement emerging from 6 studies and at the same time unsupported by 5 other studies is probably considered less tenable than a statement emerging from 6 and unsupported by 1. Some of the following 61 statements can be thought of as serious candidates for "Principles" of concept attainment, while others will be undoubtedly modified and rejected with further research.

Concept Variable

1. The conceptual rule influences ease of concept attainment; conjunctive concepts are the most easily attained while disjunctives are the most difficult. Relational and biconditional concepts are more easily attained than disjunctive concepts.[5] (10–1).[6]

2. Some primitive concepts appear to be attained more easily than others; from the easiest to most difficult, the general order is concepts of concrete objects, concepts of spatial forms, and concepts of numbers (12–4).

3. The type of concept presented may be a more significant variable in concept attainment than the method of presentation (1–0).

4. As the number of relevant properties decreases (movement from a complex to a simple concept), ease of concept attainment generally increases (12–3).

5. Exclusive concepts tend to be more easily attained than inclusive concepts (1–0).

Stimulus Variables

Nature of the Critical and Noncritical Properties

6. Ease of concept attainment does not appear to be significantly influenced by which properties become critical and which are left as noncritical (20–10).

7. As the difference between critical and noncritical properties increases, ease of concept attainment increases (3–0).

8. As critical properties become more obvious while noncritical properties become less obvious, ease of concept attainment increases (7–0).

9. As a critical property shows increasingly less variation from instance to instance, ease of concept attainment generally increases (5–0).

Amplification

9.1 If a critical property shows some variation from instance to instance, then decreasing the variation of each noncritical property on these same instances increases ease of concept attainment. If, however, no variation occurs on the critical property over the instances, then increasing the variation of each noncritical property increases ease of concept attainment (3–1).

10. For conjunctive concepts, if all the critical properties are found on a single, compact unit, concept attainment is facilitated more than if they are distributed over several units (3–0).

11. As the number of noncritical properties decreases, ease of concept attainment increases (40–1).

Amplification

11.1 Decreasing noncritical properties has a greater facilitating effect on complex concepts than on simple concepts (1–0).

11.2 Decreasing noncritical properties appears to have a facilitating effect regardless of the type of instance used (specifically, 3–2).

11.3 Decreasing noncritical properties appears to have a facilitating effect regardless of the type of concept (1–0).

Nature of the Positive and Negative Instances

12. As the difference between positive and negative instances increases, ease of concept attainment increases (7–0).

13. The degree of dissimilarity between positive and negative instances of the concept may be a more significant variable in concept attainment than the form of those instances (1–0).

14. The form of the instance apparently determines the number of noncritical properties that will be displayed or evoked by it; as this number decreases, ease of concept attainment increases. Thus, instances in verbal form—because they evoke fewer noncritical properties—increase the ease of concept attainment more than instances in visual form (e.g., drawings, photographs) (7–1).

15. A display of positive and negative instances all of which are in the same form facilitates concept attainment to a greater extent than a display which allows form variation among the instances (5–0).

16. As the degree to which negative instances fall into the same conceptual class increases, ease of concept attainment increases (2–0).

17. When instances are in verbal form: As the degree to which the positive instances tend to evoke the appropriate critical properties (conceptual responses) and negative instances tend to evoke the appropriate noncritical properties increases, ease of concept attainment increases (11–0).

Subject Variables

Age

18. As age increases, ease of concept attainment increases (16–3).

19. Age appears to influence the selection of critical properties: Below 3 years, form is domi-

---

[5] See Section V for research pertaining to this statement.

[6] The first number in parentheses after each statement represents the total number of studies from which the statement was derived; the second number represents the total number of studies which do not support the statement.

nant; from 3 to 6 years, color becomes dominant; from 6 years on, form again becomes dominant. As age increases, dominance of physical characteristics seems to give way to dominance of functional characteristics (10–0).

Intelligence

20. As intelligence increases, ease of concept attainment generally increases (8–4).

Sex

21. The sex of the student does not appear to be a significant variable in the ease with which concepts are attained (10–4).

Personality Factors

22. Anxiety probably interacts with other variables (e.g., task complexity, intelligence) in influencing ease of concept attainment.

Amplification

22.1 As anxiety increases ease of concept attainment increases in general (2–1).

22.2 As anxiety increases ease of concept attainment increases among higher IQ students (1–0).

22.3 As anxiety increases ease of concept attainment decreases with more complex concepts (2–2).

23. As rigidity decreases, ease of shifting from the learning of one concept to the learning of another concept may increase (1–0).

24. As the degree of ego involvement increases, ease of concept attainment may increase on more complex concepts (1–0).

25. If the student has a natural preference for one property over another, ease of concept attainment increases if the preferred property is critical and decreases if it is noncritical (2–0).

Task Variables

Instructions

26. Directions given to the student which focus him on the critical and/or noncritical properties and/or conceptual rule increase the ease with which he will attain the concept more than nonfocusing or no directions (8–1).

27. Pointing out the critical properties and/or the conceptual rule increases ease of concept attainment more than permitting the student to discover them (7–1).

Preinstance Presentation Activities

28. As simple prior exposure to labels (of critical properties, instances, concepts), critical properties, and instances increases, ease of subsequent concept attainment increases (13–0).

29. Certain pretask associational activities increase the ease with which subsequent concepts are attained.

Amplification

29.1 Associating each critical property with its label increases ease of subsequent concept attainment (2–0).

29.2 Associating the concept label with (*a*) each critical property or its label, (*b*) each concept instance increases ease of subsequent concept attainment (6–0).

Instance Presentation

30. An orderly or systematic display of instances increases ease of concept attainment more than a disorderly or random display (1–1).

31. Presenting instances simultaneously and allowing them to remain in full view during the concept task increases the ease of concept attainment more than a successive-withdrawal presentation (11–2).

32. For the greatest ease of concept attainment, the optimal number of instances presented simultaneously appears to be about four; as fewer or greater numbers are used, ease of concept attainment decreases (4–1).

33. Allowing ample instance exposure time increases ease of concept attainment (3–0).

34. Generally, as the proportion of positive instances increases, ease of concept attainment increases (6–2).

Amplification

34.1 For conjunctive concepts the presentation sequence from most to least effective is as follows: (*a*) a sequence of all positive instances, (*b*) a sequence of several (more than two) positive instances followed by a sequence of several negative instances, (*c*) a sequence of mixing positive and negative instances, (*d*) a sequence of all negative instances (20–5).

Qualification

34.2 For disjunctive concepts, a sequence of negative instances probably increases the ease of concept attainment more than a sequence of positive instances (4–0).

35. As critical properties become less obvious, repetition of the same positive instances increases the ease of concept attainment more than nonrepetition. As they become more obvious, repetition becomes less important (1–0).

36. As the number of instances presented increases, retention of the concept increases (1–0).

37. Positive instances which appear in an obvious to subtle sequence increase the ease of concept attainment more than when in a subtle to obvious sequence (1–0).

Student Responses

38. Ease of concept attainment and transfer to new concepts occurs from greatest to least when students (*a*) verbalize critical properties only, (*b*) verbalize critical and noncritical properties, (*c*) do not verbalize, (*d*) verbalize noncritical properties only (4–1).

39. Ease of concept attainment and degree of retention increase as the student is able to (*a*) verbalize both the critical properties and the con-

cept name, (*b*) verbalize the concept name throughout the concept task (2–0).

Qualification

39.1 A student can have acquired a concept (correctly sort and correctly apply in another situation) and yet be unable to verbalize the critical properties and conceptual rule (2–0).

40. When a student gives an incorrect response, having him give the correct response increases ease of concept attainment more than simply informing him of the correct response (1–0).

41. As the number of response categories open to the student decreases (e.g., from three categories to example or nonexample) ease of concept attainment increases (1–0).

42. Ease of concept attainment increases more if the student is permitted to physically manipulate the instances than if he is not allowed to do so (1–0).

Feedback

43. Ease of concept attainment and degree of retention are increased under the following conditions: any feedback is more facilitating than no feedback; feedback for (*a*) incorrect responses only or (*b*) correct and incorrect responses is more facilitating than feedback for correct responses only; there are no apparent differences in facilitation between feedback for incorrect responses only and feedback for correct and incorrect responses (9–2).

44. As the percentage of feedback to concept responses increases, ease of concept attainment increases (100%—feedback after every response to each instance) (9–0).

Amplification

44.1 Increasing the percentage of feedback becomes more important as concepts become more complex (1–0).

Qualification

44.2 As the percentage of feedback to concept responses decreases, retention of the concept increases (2–0).

45. As the degree of completeness and precision of feedback increase, ease of concept attainment increases (4–0).

46. As the percentage of irregularity of misinformation feedback to concept responses decreases, ease of concept attainment increases (8–0).

Amplification

46.1 Decreasing percentage of misinformation becomes more important for complex concepts (3–0).

47. Ease of concept attainment increases more if the feedback occurs in the presence of the instances under consideration than if the instances are not present (1–0).

48. When determining the degree to which a student has attained a given concept, allowing him to compare his correct and incorrect identifications of new positive and negative instances of the concept with the original display of concept instances increases ease of concept attainment more than if he is not allowed such a comparison (1–0).

49. As delay of feedback to concept responses decreases, ease of concept attainment increases (2–1).

50. As the postinformation-feedback interval (time between feedback to student and presentation of next instance) increases, ease of concept attainment increases (2–0).

Amplification

50.1 Increasing this interval becomes more important for complex concepts (1–0).

51. Generally, as the intensity of feedback decreases, ease of concept attainment increases (1–0).

52. An impersonal source which provides feedback for incorrect responses may increase the ease of concept attainment more than when feedback is from a personal source (teacher) (1–0).

53. The same type of feedback may be effective over several age levels (4–9 years) (1–0).

Practice

54. Practice in attaining concepts increases the ease with which new concepts are attained (14–0).

Other Variables within the Learning Task

55. The order of presenting conjunctive and disjunctive concepts does not appear to influence the ease with which they are attained (1–0).

56. As the difference between two concepts presented simultaneously increases, the ease of concept attainment increases (2–0).

57. If a compound concept is to be learned, complete learning of the component concepts increases the ease with which the compound concept is attained more than does incomplete learning (4–0).

58. Students tend to imitate the concept performance of others participating in the same concept task (2–0).

59. Students who are not distracted by watching other students' concept performance attain concepts more easily and quickly than if distracted (1–0).

60. Two students working on the same concept attain it more easily than if they were working separately (3–0).

Amplification

60.1 The effect of two students working together is more important with complex concepts than with simple concepts (1–0).

61. Greatest transfer to (*a*) new instances of the same concept, and (*b*) new but related concepts occurs when a variety of original instances (concepts) is learned and each of these original instances (concepts) is completely learned (10–0).

SECTION IV. INSTRUCTIONAL PRESCRIPTIONS FOR TEACHING CONCEPTS SUGGESTED BY THE STATEMENTS ABOUT CONCEPT ATTAINMENT

The 61 statements about concept attainment are descriptive in nature and do not directly prescribe teaching practices. Therefore, it seems unlikely that a teacher, planning to teach a specific concept, would view them as helpful in their present form. Indeed, few teachers would possess the time or energy to tease out prescriptions for lesson plans from such an overpowering and technical list. As one begins to perform this function for the teacher, it becomes quickly apparent that it is in this task that the greatest amount of interpretation and data reduction occurs. Unless considerable care is taken, the precision that can be generally maintained in moving from research results to statements about concept attainment is rapidly dissipated in moving from statements to prescriptions for teaching.

In an attempt to minimize loss of precision, the decision was made to retain much of the technical language in developing teaching prescriptions from the statements. Whether or not a teacher could even now find these helpful depends on her understanding the definitions of the two main parameters (see definitions below).

The teasing-out task consisted of the following. A general model of teaching practice was outlined for the purpose of categorizing the 61 statements. It consisted of the following five steps or phases a teacher might go through in preparing and teaching: (1) selection of the concept to be taught and a conceptual objective from the examples in Section II (e.g., The student will have formed the concept "square"); (2) preparation of materials for teaching the concept; (3) preinstruction evaluation; (4) teaching the concept; (5) postinstruction evaluation.

If a statement was considered relevant to one of the phases of the model, it was placed in that category. For example, Statement 15 indicates that when all instances of a concept are of the same form or type, concept attainment is facilitated. This statement was considered relevant to Phase 2, preparation of materials. Thirty-seven of the statements, or approximately 61 %, could be comfortably placed in the five phases. Statements like Number 21, "The sex of the student does not appear to be a significant variable in the ease with which concepts are attained," did not appear immediately relevant to any phase and were omitted.

Statements within a phase were either combined or left separately and translated into a teaching prescription. In each case, an attempt was made to tease out teaching prescriptions which were as specific as possible even though several statements went into one prescription. For example, in Phase 4, Statements 40, 43, 44, 45, 46, 47, 48, and 49 were translated into the following prescription: "After either a correct or incorrect response to *each* instance, give the student immediate, complete, and accurate feedback. After feedback for an incorrect response, make certain the student verbalizes the correct response (Phase 4, 13.3)." Often, the statements themselves would suggest a sequence of prescriptions within a phase (Phase 4, for example).

Following are some 16 teaching prescriptions throughout the five phases of the teaching model, prefaced by a set of definitions. In order to observe how each of these prescriptions can be applied to an actual situation, examples for teaching the concept "square" are provided.[7]

The general question of whether a teacher using these experimentally derived prescriptions will, in fact, effect greater concept attainment than a teacher using nonexperimentally derived practices demonstrably different in nature is currently under exploration by the author and his associates.

Instructional Prescriptions

Definitions

1. A *critical property* is a necessary characteristic of a given concept. A particular set of critical properties in their appropriate relationship constitutes a concept.

2. A *noncritical property* is an unnecessary characteristic of a given concept.

3. A *positive instance* is a unit which contains or displays all the critical properties of a concept in their appropriate relationship.

---

[7] These teaching prescriptions are more easily used when the instances are of a concrete nature (e.g., observable objects). As the instances become more abstract (e.g., observable events, partly observable processes, words) difficulty arises in identifying the properties, if not the instances. However, much of the difficulty may be simply due to inexperience in seeing the properties and instances of a more abstract concept.

There are two types described below, positive introductory and positive confirmatory.

4. A *negative instance* is a unit which either contains or displays (*a*) none or some but not all the critical properties of a concept in their appropriate or inappropriate relationship, or (*b*) all the critical properties but in an inappropriate relationship.

Phase 1: Selection of Concept and Conceptual Objectives

1. Select the particular concept to be taught and place it into one or more of the conceptual objectives in Section II.

2. Decide upon the concept's specific critical properties as well as its frequently observed noncritical properties.

3. Determine prerequisite concepts and performance skills and make certain they are in the student's repertory (57[8]).

Phase 2: Preparation of Materials

4. There are three types of instances to be obtained or constructed by the teacher. Every instance in all three types should be presented or exposed as a single, compact unit (10), and all instances should be of the same general form (15).

5. *Positive-introductory instances.* Obtain instances which are totally free of noncritical properties, or at least those which contain or display as few as possible (11). If some noncritical properties are unavoidable on a given instance, try to change the instances so that (*a*) these noncritical properties are as dissimilar as possible from the critical properties (7); (*b*) these noncritical properties are less obvious and the critical properties are more obvious (8); (*c*) a noncritical property shows as much variation as possible from instance to instance, while a critical property shows as little variation as possible (9, 9.1).

6. *Positive-confirmatory instances.* Obtain other instances each of which contains or displays (*a*) all of the critical properties, and (*b*) at least one and preferably several noncritical properties. Also, a critical property in this group should show some variation over the instances; likewise, with a noncritical property. (This variation need not occur with every critical and every noncritical property from instance to instance.)

7. *Negative instances.* Obtain other instances each of which contains or displays (*a*) either an incomplete set of critical properties (ranging from none of the critical properties to most of the critical properties) or a complete set in an inappropriate relationship, and (*b*) several noncritical properties. Some of these noncritical properties should be identical to those contained by the positive-confirmatory instances, and some should be different (12, 13). These negative instances should be, as much as possible, all from the same conceptual class. In general, they are to be mostly similar to each other and mostly different from both types of positive instances (12, 13, 16).

Phase 3: Preinstruction Evaluation

8. Refer to the conceptual objectives that were selected during Phase 1, Number 1.

9. Identify their learning products in column 3 of Table 1.

10. Using the concept to be taught, develop a pretest using the appropriate formats in column 3.

Phase 4: Teaching the Concept

11. There are several procedures that should be followed throughout both the introductory and confirmatory stages of teaching the concept:

11.1. Present neither too many nor too few instances. The appropriate number will vary according to age of the student and the concept being presented (32).

11.2. Always present the instances in an orderly or systematic manner (30).

11.3. For a given stage, present all the instances simultaneously; allow them to remain in full view throughout both stages (31).

11.4. Allow the student ample exposure time with the instances (33).

11.5. Review of the instances presented may at times be necessary and is always helpful (35, 61).

---

[8] Statement(s) in Section III which suggests this teaching prescription. It can be noted that the teaching prescriptions have been generated entirely from statements concerning stimulus variables and statements concerning task variables. This is because the teacher has little control over the type of concept to be taught (concept variables) or the student's characteristics (subject variables).

11.6. When feasible, allow the student access to the instances (42).

12. Introductory stage:

12.1. Present only positive-introductory instances, naming each as the concept (28, 29.2, 34, 34.1, 37).

12.2. On each instance, point out and name all the critical properties and their appropriate combination (26, 27, 28, 29.1). Also, point out any noncritical properties which might be present (26, 27).

13. Confirmatory stage:

13.1. Next, present a mixture of positive-confirmatory and negative instances (34.1, 36, 37, 54, 61).

13.2. Have the student respond to each instance as either an example of the concept or a nonexample of the concept, and have him verbalize the critical properties and concept name whenever possible (38, 39, 41).

13.3. After either a correct or incorrect response to *each* instance, give the student immediate, complete, and accurate feedback. After feedback for an incorrect response, make certain the student verbalizes the correct response (40, 43, 44, 45, 46, 47, 48, 49).

13.4. Allow ample time between feedback to the student's responses and teacher's focus on the next instance (48, 50).

13.5. When possible, present additional positive-confirmatory and negative instances which show increasingly more variation on both critical and noncritical properties. Continue to follow the same procedure outlined for the confirmatory stage (36, 37, 54, 61).

Phase 5: Postinstruction Evaluation

14. Refer to the previously selected learning products in column 3 of Table 1.

15. Using the concept taught, develop a posttest (parallel form of the pretest) using the appropriate formats in column 3.

16. Decide upon what degree of change between pretest and posttest performance will constitute "concept formation"; or decide at what level the posttest performance must be to represent "concept formation."

Example of Instructional Prescriptions for the Concept "Square"

Phase 1: Selection of Concept and Conceptual Objectives

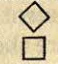
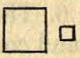
1. During the learning process the student will have formed the concept "square" (Number 2, Table 1).

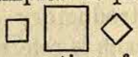
2. Critical properties are as follows: (*a*) All squares have four sides. (*b*) All the sides of a square are equal in length. (*c*) All the corners of a square are right angles. Frequently observed noncritical properties are as follows: (*a*) How a square is oriented in space is not critical. (*b*) The size of a square is not critical.

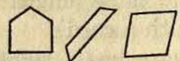
3. Prerequisite concepts are (*a*) right angle, (*b*) four, (*c*) side, (*d*) equal, (*e*) length, (*f*) corners. Prerequisite skills are (*a*) how to measure a line with a card, (*b*) how to measure an angle with the card's corner.

Phase 2: Preparation of Materials

4. An instance is always contained in a single unit. For example: □ All instances are made from paper. (There are no cloth instances nor square-object instances included).

5. Examples of positive-introductory instances: □ □ (Not  or )

6. Examples of positive-confirmatory instances:  (The nature of the critical properties of this particular concept precludes much variation of the critical properties.)

7. Examples of negative instances:  (Also, the nature of this particular concept tends to preclude a complete set of critical properties in an inappropriate relationship; each of these negative instances has an incomplete set of critical properties.)

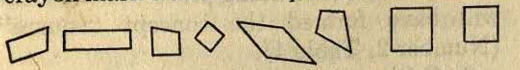
Phase 3: Preinstruction Evaluation

8. During the learning process the student will have formed the concept "square" (Number 2, Table 1).

9. As a learning product, the student will be able to, when presented with a group of

instances, correctly sort the examples from the nonexamples (Number 2, Table 1).

10. Pretest: Instances similar to those below are all presented on one sheet of paper. The student might be asked to place a crayon mark on all the "squares" he can find.



Phase 4: Teaching the Concept

11. There are several procedures that should be followed throughout both the introductory and confirmatory stages of teaching the concept:

11.1. With preschool children, the teacher might first try one or two positive-introductory and from two to four positive-confirmatory and negative instances.

11.2. The instances are cut out from white paper and are arrayed in an orderly manner on a flannel board.

11.3. All positive-introductory are presented simultaneously as are all positive-confirmatory and negative; each is left on the flannel board.

11.4. Exposure time for each instance depends upon the teacher's judgment and initial student responses.

11.5. Review of instances occurs whenever appropriate throughout the task.

11.6. The student is allowed to manipulate and compare the paper instances.

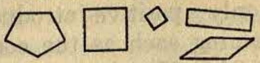
12. Introductory stage:

12.1. Positive introductory instances are first presented; each is referred to as a "square."

12.2. "See how this square has four sides" (Teacher counts the sides). "The sides of this square are the same length" (She measures each side with a card). "Every corner of this square has a right angle" (She fits the corner of the card into each of the corners of the square). (There are no critical properties). This procedure is carried out on each of the positive-introductory instances.

13. Confirmatory stage:

13.1. Examples of positive-confirmatory and negative instances are as follows:



13.2. The teacher points to each instance and says, "Use your card and tell me whether or not this is a square." "What does it have to have to be a square?"

13.3. After each response to each instance, the teacher gives complete feedback. For example: "Yes, you are right. This is a square even though it is tipping over, because it has. . . ." Or, "I don't think this is a square because the sides don't all seem to be the same length. Why don't you measure it again. Are all the sides of a square supposed to be the same length?"

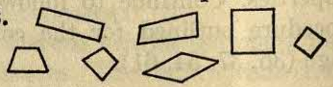
13.4. The teacher allows ample time for the student to study the instance after her feedback to his response and before going on to the next one.

13.5. With this particular concept, additional positive-confirmatory and negative instances may not be required.

Phase 5: Postinstruction Evaluation

14. As a learning product, the student will be able to, when presented with a group of instances, correctly sort the examples from the nonexamples (Number 2, Table 1).

15. Posttest: Instances similar to those below are presented on one sheet of paper. The student is asked to mark all the "squares." This would be a parallel form of the pretest.



16. If the student can correctly sort—by marking—90% of all the instances, this would be taken as evidence that he has formed the concept "square." (Under other circumstances, a performance gain from the pretest to the posttest may be desired.)

SECTION V. RESEARCH PERTAINING TO EACH STATEMENT

Statement	*Research from which statement was derived*	*Research not supporting the statement*
1	Braley (1963) Conant & Trabasso (1964) Haygood & Stevenson (1967) Hunt & Hovland (1960)	Haygood and Devine (1967)

SECTION V—*Continued*

Statement	*Research from which statement was derived*	*Research not supporting the statement*
	Laughlin & Jordon (1967) Nicholson (1966) Schwartz (1966) Wallace & Sechrest (1961) Wells & Deffenbacher (1967) Youniss & Furth (1967)	
2	Deterline (1957) Heidbreder (1946) Heidbreder (1947) Heidbreder (1948) Heidbreder (1949a) Heidbreder (1949b) Heidbreder, Bensley, & Ivy (1948) Heidbreder & Overstreet (1948) Jeffrey & Kluppel (1962) Juzak (1955) Komachiya (1957) Wenzel & Flurry (1948)	Dattman & Israel (1951) Grant & Curran (1952) Grant, Jones, & Tallantis (1949) Nadelman (1957)
3	Schwartz (1966)	
4	Bourne (1963b) Bourne (1965) Bourne, Goldstein, & Link (1964) Bourne & Haygood (1959) Bulgarell & Archer (1962) Glanzer, Huttenlocher, & Clark (1963) Haygood (1965) Kepros & Bourne (1966) Krossner (1966) Schvaneveldt (1966) Shepard, Hovland, & Jenkins (1961) Walker & Bourne (1961) Wallach (1962)	Laughlin (1966) Laughlin & Jordon (1967) Laughlin & Doherty (1967)
5	Osler & Fivel (1961)	
6	Archer, Bourne, & Brown (1955) Bourne (1963a) Bourne (1957) Bourne (1965) Gelfand (1958) Haygood & Stevenson (1957) Johnson (1966) Katz (1968) Keele & Archer (1967) Laughlin (1966) Laughlin (1967) Laughlin & Doherty (1967) Namikas (1967) Osler & Kofsky (1965) Pishkin (1960) Pishkin (1961a) Pishkin (1961b) Pishkin (1965) Walker & Bourne (1961) Wolfgang, Pishkin, & Lundy (1962)	Archer (1962) Bourne (1963b) Bourne & Pendelton (1958) Kepros & Bourne (1966) Pishkin & Shurley (1965) Rasmussen & Archer (1961) Richrdson & Bergum (1954) Trabasso (1963) Trabasso (1967) Wallace (1964)
7	Dietz (1955) Rankin (1954) Sechrest & Kaas (1965)	

SECTION V—*Continued*

Statement	*Research from which statement was derived*	*Research not supporting the statement*
8	Archer (1962) Dattman & Israel (1951) Heidbreder (1949) Heidbreder & Zimmerman (1955) Runquist & Hutt (1961) Shore & Sechrest (1961) Trabasso (1963)	
9	Battig & Bourne (1961) Cahill & Hovland (1960) Dick (1963) Morrisett & Hovland (1959) Saltz & Sigel (1967)	
9.1	Detambel & Stolurow (1956) Saltz & Sigel (1967) Trabasso (1963)	Anderson & Guthrie (1966)
10	Bourne & Parker (1964) Wallach (1962) Wells (1965)	
11	Anderson (1966) Archer (1962) Archer, Bourne, & Brown (1955) Baggaley, Havas, & Stanners (1959) Battig & Bourne (1961) Bourne (1957) Bourne & Bunderson (1963) Bourne, Guy, Dodd, & Justesen (1965) Bourne & Haygood (1959) Bourne & Haygood (1960) Bourne & Pendleton (1958) Braley (1962) Braley & Johnson (1963) Brown & Archer (1956) Bulgarella & Archer (1962) Gelfand (1958) Haygood & Bourne (1964) Haygood & Devine (1967) Haygood & Stevenson (1967) Johannsen (1962) Johnson (1966) Johnson & O'Reily (1964) Katz (1968) Keele & Archer (1967) Kepros & Bourne (1966) Kirloskar & Parameswaran (1967) Lordahl (1961) Meyer & Offenbach (1962) Osler & Kofsky (1965) Osler & Kofsky (1966) Peterson (1962) Pishkin (1960) Pishkin (1961a) Pishkin (1961b) Pishkin (1965) Rasmussen & Archer (1961) Scholl (1966) Walker & Bourne (1961) Wolfgang (1967) Wolfgang, Pishkin, & Lundy (1962)	Wallace & Sechrest (1961)

SECTION V—*Continued*

Statement	*Research from which statement was derived*	*Research not supporting the statement*
11.1	Haygood & Stevenson (1967)	
11.2	Bulgarella & Archer (1962) Keele & Archer (1967) Pishkin & Shurley (1965)	Bourne (1965) Lordahl (1961)
11.3	Haygood & Stevenson (1967)	
12	Baum (1954) Deno (1965) Newman (1956) Rankin (1954) Sechrest & Kaas (1965) Spiker (1963) Underwood (1957)	
13	Deno (1965)	
14	Baggaley (1956) Davidson (1952) Johnson & O'Reilly (1964) Nicholson (1966) Osler & Trautman (1961) Runquist & Hutt (1961) Wolff (1967)	Heidbreder (1949)
15	Davidson (1952) Haygood (1965) Lordahl (1961) Otto (1962) Otto (1964)	
16	Duncan (1965) Kendler & Karasik (1959)	
17	Bousfield & Ruff (1964) Coleman (1964) Crouse & Duncan (1963) Freedman & Mednick (1958) Kendler & Karasik (1959) Mayzner & Tresselt (1962) Mednick & Freedman (1960) Mednick & Halpern (1962) Pollack (1963) Schulz, Miller, & Radtke (1963) Underwood & Richardson (1956)	
18	Katz (1968) Long (1940) Long & Welch (1941) Odom (1966) Osler & Fivel (1961) Osler & Kofsky (1965) Osler & Shapiro (1964) Osler & Trautman (1961) Osler & Weiss (1962) Piaget (1930) Pishkin, Wolfgang, & Rasmussen (1967) Runquist & Hutt (1961) Saltz & Sigel (1967)	Friedman (1965) Klugh & Roehl (1965) Meyer & Seidman (1961)

SECTION V—*Continued*

Statement	*Research from which statement was derived*	*Research not supporting the statement*
	Sigel (1953) Wylie (1958) Yudin & Kates (1963)	
19	Birch & Bortner (1966) Brian & Goodenough (1929) Colby & Robertson (1942) Corah (1964) Kagan (1961) Kendler (1961) Lee (1965) Russell & Saadeh (1962) Suchman (1966) Wylie (1958)	
20	Beilin (1966) Denny (1966) Katz (1968) Nicholson (1966) Osler & Fivel (1961) Osler & Weiss (1962) Weiner (1967) Wolff (1967)	Braley (1962) Osler & Shapiro (1964) Osler & Trautman (1961) Wylie (1958)
21	Beilin (1966) Conant & Trabasso (1964) Nadelman (1957) Olson (1963) Osler & Kofsky (1965) Pishkin (1960) Pishkin (1967) Pishkin, Wolfgang, & Rasmussen (1967) Rasmussen & Archer (1961) Schvaneveldt (1966)	Archer (1962) Forgua & Fowler (1957) Pishkin & Blanchard (1964) Pishkin & Shurley (1965)
22.1	Wesley (1953) Wolfgang, Pishkin, & Lundy (1962)	Braley (1962)
22.2	Denny (1966)	
22.3	Holbrook (1954) Romanow (1958)	Arnhoff (1959) Wolfgang, Pishkin, & Lundy (1962)
23	Wesley (1953)	
24	Romanow (1958)	
25	Pishkin (1961a) Suchman & Trabasso (1966)	
26	Archer, Bourne, & Brown (1955) McCreary (1963) McGuigan (1958) Osler & Weiss (1962) Pishkin (1965) Saunders & Neimark (1967) Underwood & Richardson (1956) Wittrock (1963)	Sechrest & Wallace (1962)
27	Goss & Moylan (1958) Haygood & Divina (1967) Haygood & Stevenson (1967) Klausmeier & Meinke (1968)	Della-Piana (1957)

SECTION V—*Continued*

Statement	*Research from which statement was derived*	*Research not supporting the statement*
	McCreary (1963) Underwood & Richardson (1956) Weiner (1967)	
28	Cantor (1955) Carey & Goss (1957) Forgua & Fowler (1957) Goss & Moylan (1958) Friffin & Beier (1961) Haygood (1965) Kluppel & Jeffrey (1964) Lordahl (1961) Norcross & Spiker (1957) Shepard & Schaeffer (1956) Spiker (1956) Stern (1965) Weir & Stevenson (1959)	
29.1	Gelfand (1958) Namikas (1967)	
29.2	Carey & Goss (1957) Fenn & Goss (1957) Jeffrey & Kluppel (1962) Rasmussen & Archer (1961) Thysell & Schulz (1964) Wolff (1967)	
30	Grant (1951)	Laughlin (1965)
31	Bourne (1963b) Bourne, Goldstein, & Link (1964) Bourne, Guy, & Justesen (1965) Cahill & Hovland (1960) Crouse & Duncan (1963) Hunt (1961) Kates & Yudin (1964) Reed (1950) Scholl (1966) Schwartz (1966) Weiner (1967)	Huttenlocher (1962a) Nadelman (1957)
32	Bourne, Goldstein, & Link (1964) Pishkin & Wolfgang (1965) Reed (1946) Reed (1950)	Pishkin (1967)
33	Bourne, Goldstein, & Link (1964) Glanzer, Huttenlocher, & Clark (1963) Green (1955)	
34	Haygood & Devine (1967) Haygood & Stevenson (1967) Long (1965) Mayzner (1962) Reed (1946) Schvaneveldt (1966)	Buss (1950) Shore & Sechrest (1961)
34.1	(a) was more effective than: (b) Schulz, Miller, & Radtke (1963) (c) Bourne & Jennings (1963) Fryatt & Tulving (1963) Glanzer, Huttenlocher, & Clark (1963) Griffith, Spitz, & Lipman (1959)	Whitman & Garner (1963) Anderson & Guthrie (1966) Olson (1963) Smoke (1932) Smoke (1933)

SECTION—V *Continued*

Statement	*Research from which statement was derived*	*Research not supporting the statement*
	Hovland & Weiss (1953) Newman (1956) Peterson (1962) Restle & Emmerich (1966) Whitman & Garner (1963) (d) Freibergs & Tulving (1961) (b) was more effective than: (c) Deno (1965) Glanzer, Huttenlocher, & Clark (1963) Kurtz & Hovland (1956) Peterson (1962) Schulz, Miller, & Radtke (1963) Whitman & Garner (1963) (c) was more effective than: (d) Hovland & Weiss (1953) Huttenlocher (1962b) Smoke (1932)	
34.2	Bruner, Goodenough, & Austin (1956) Conaut (1966) Wallace & Sechrest (1961) Wells & Watson (1965)	
35	Shore & Sechrest (1961)	
36	Erlebacher & Archer (1961)	
37	Olson (1963)	
38	Kendler & Kendler (1961) Nilgram & Noce (1968) Namikas (1967) Schwartz (1948)	Kendler, Kendler, & Wells (1960)
39	Reed (1946) Wolff (1967)	
39.1	Furth (1961) Schwartz (1948)	
40	Suppes & Ginsberg (1962)	
41	Pollack (1963)	
42	Davidson (1952)	
43	Brackbill & O'Hara (1958) Buss, Braden, Orgel, & Buss (1956) Buss & Buss (1956) Jazak (1955) Meyer & Offenbach (1962) Meyer & Seidman (1960) Meyer & Seidman (1961) Penny & Lupton (1961) Sullivan (1960)	Pishkin (1967) Sullivan (1960)
44	Bourne & Haygood (1960) Bourne & Pendleton (1958) Carpenter (1953) Erlebacher & Archer (1961) Green (1955) Meyer & Offenbach (1962) Namikas (1967) O'Connell (1965) Osler & Shapiro (1964)	

SECTION V—*Continued*

Statement	*Research from which statement was derived*	*Reseach not supporting the statement*
44.1	Meyer & Offenbach (1962)	
44.2	Green (1955) O'Connell (1965)	
45	Bourne & Haygood (1960) Bourne & Pendleton (1958) Meyer & Seidman (1961) O'Connell (1965)	
46	Bourne (1963a) Johannsen (1962) Morin (1955) Pishkin (1960) Pishkin (1961b) Pishkin (1965) Saunders & Neimark (1967) Wolfgang, Pishkin, & Lundy (1962)	
46.1	Pishkin (1960) Pishkin (1961b) Wolfgang, Pishkin, and Lundy (1962)	
47	Restle & Emmerich (1966)	
48	Pishkin & Wolfgang (1965)	
49	Bourne (1957) Sax (1960)	Bourne & Bunderson (1963)
50	Bourne & Bunderson (1963) Bourne, Guy, & Justeson (1965)	
50.1	Bourne & Bunderson (1963)	
51	Wallace (1964)	
52	Wallace (1964)	
53	Meyer & Seidman (1961)	
54	Bourne (1963a) Brown & Archer (1956) Buss (1950) Conant (1966) Freibergs & Tulving (1961) Klausmeier & Mainke (1968) Morrisett & Hovland (1959) Olson (1963) Pishkin & Shurley (1965) Semme & Warren (1955) Scholl (1966) Schulz, Miller, & Radtke (1963) Tavrow (1965) Wells & Watson (1965)	
55	Conant & Trabasso (1964)	
56	Newman (1956) Underwood (1957)	
57	Grant & Berg (1948) Grant & Cost (1954) Kendler & Vineberg (1954) Morrisett & Hovland (1959)	

Section V—*Continued*

Statement	*Research from which statement was derived*	*Research not supporting the statement*
58	Allen & Bragg (1968) Pishkin & Blanchard (1963)	
59	Wolfgang (1967)	
60	Laughlin (1965) Laughlin & Doherty (1967) Wolfgang (1967)	
60.1	Wolfgang (1967)	
61	Adams (1954) Callertine & Warren (1955) Caul & Ludvigson (1964) Dick (1963) Grant & Berg (1948) Grant & Cost (1954) Morrisett & Hovland (1959) Semme & Warren (1955) Stern (1965) Youniss & Furth (1967)	

REFERENCES

Adams, J. A. Multiple versus single problem training in human problem solving. *Journal of Experimental Psychology,* 1954, **48,** 15–19.

Allen, V. L., & Bragg, B. W. Effect of social pressure on concept identification. *Journal of Educational Psychology,* 1968, **59,** 302–308.

Anderson, R. C. Sequence constraints and concept identification. *Psychological Reports,* 1966, **19,** 1295–1302.

Anderson, R. C., & Guthrie, J. T. Effects of some sequential manipulations of relevant and irrelevant stimulus dimensions on concept learning. *Journal of Experimental Psychology,* 1966, **72,** 501–504.

Archer, E. J. Concept identification as a function of obviousness of relevant and irrelevant information. *Journal of Experimental Psychology,* 1962, **63,** 616–620.

Archer, E. J., Bourne, L. E., Jr., & Brown, F. G. Concept identification as a function of irrelevant information and instructions. *Journal of Experimental Psychology,* 1955, **49,** 153–164.

Arnhoff, F. N. Stimulus generalization and anxiety. *Journal of General Psychology,* 1959, **60,** 131–136.

Baggaley, A. R. Concept learning by college students as a function of personalization and structural variables. *American Psychologist,* 1956, **11,** 445.

Baggaley, A. R., Havas, F. W., & Stanners, R. F. Effects of stimulus complexity and discriminability on concept learning. *Psychological Reports,* 1959, **5,** 757–763.

Battig, W. F., & Bourne, L. E., Jr. Concept identification as a function of intra- and interdimensional variability. *Journal of Experimental Psychology,* 1961, **61,** 329–333.

Baum, M. H. Simple concept learning as a function of intralist generalization. *Journal of Experimental Psychology,* 1954, **47,** 89–94.

Beilin, H. Feedback and infralogical strategies in invariant area conceptualization. *Journal of Experimental Child Psychology,* 1966, **3,** 267–278.

Birch, H. G., & Bortner, M. Stimulus competition and category usage in normal children. *Journal of Genetic Psychology,* 1966, **109,** 195–204.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.). *Taxonomy of educational objectives: Cognitive domain.* New York: David McKay, 1956.

Bourne, L. E., Jr. Effects of delay of information feedback and task complexity as the identification of concepts. *Journal of Experimental Psychology,* 1957, **54,** 201–207.

Bourne, L. E., Jr. Factors affecting strategies used in problems of concept formation. *American Journal of Psychology,* 1963, **76,** 229–238. (a)

Bourne, L. E., Jr. Long-term effects of misinformative feedback upon concept identification. *Journal of Experimental Psychology,* 1963, **65,** 137–147. (b)

Bourne, L. E., Jr. Hypotheses and hypothesis shifts in classification learning. *Journal of General Psychology,* 1965, **72,** 251–261.

Bourne, L. E., Jr., & Bunderson, C. V. Effects of delay of informative feedback and length of postfeedback interval on concept identification. *Journal of Experimental Psychology,* 1963, **65,** 1–5.

Bourne, L. E., Jr., Goldstein, S., & Link, W. E. Concept learning as a function of availability of previously presented information. *Journal of Experimental Psychology,* 1964, **67,** 439–448.

Bourne, L. E., Jr., Guy, D. E., Dodd, D. H., & Justesen, D. R. Concept identification: The effects of varying length and informational components of the intertrial interval. *Journal of Experimental Psychology,* 1965, **69,** 624–629.

Bourne, L. E., Jr., & Haygood, R. C. The role of stimulus redundancy in concept identification.

Journal of Experimental Psychology, 1959, **58,** 232–238.

BOURNE, L. E., JR., & HAYGOOD, R. C. Effects of intermittent reinforcement as an irrelevant dimension and task complexity upon concept identification. *Journal of Experimental Psychology,* 1960, **60,** 371–375.

BOURNE, L. E., JR., & JENNINGS, P. The relationship between response contiguity and classification learning. *Journal of General Psychology,* 1963, **69,** 335–338.

BOURNE, L. E., JR., & PARKER, B. K. Differences among modes for portraying stimulus information concept identification. *Psychonomic Science,* 1964, **1,** 209–210.

BOURNE, L. E., JR., & PENDLETON, R. B. Concept identification as a function of completeness and probability of information feedback. *Journal of Experimental Psychology,* 1958, **56,** 413–419.

BOUSFIELD, W. A., & RUFF, C. R. Clustering as a function of response dominance. *Journal of Experimental Psychology,* 1964, **67,** 76–79.

BRACKBILL, Y., & O'HARA, J. The relative effectiveness of reward and punishment for discrimination learning in children. *Journal of Comparative and Physiological Psychology,* 1958, **51,** 747–751.

BRALEY, L. S. Some conditions influencing the acquisition and utilization of cues. *Journal of Experimental Psychology,* 1962, **64,** 62–66.

BRALEY, L. S. Strategy selection and negative instances in concept learning. *Journal of Educational Psychology,* 1963, **54,** 154–159.

BRALEY, L. S., & JOHNSON, D. M. Novelty effects in cue acquisition and utilization. *Journal of Experimental Psychology,* 1963, **66,** 421–422.

BRIAN, C., & GOODENOUGH, F. L. The relative potency of color and form perception at various ages. *Journal of Experimental Psychology,* 1929, **12,** 197–213.

BROWN, F. G., & ARCHER, E. J. Concept identification as a function of task complexity and distribution of practice. *Journal of Experimental Psychology,* 1956, **52,** 316–321.

BRUNER, J. C., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

BULGARELLA, R. G., & ARCHER, E. J. Concept identification of auditory stimuli as a function of amount of relevant and irrelevant information. *Journal of Experimental Psychology,* 1962, **63,** 254–257.

BUSS, A. H. A study of concept formation as a function of reinforcement and stimulus generalization. *Journal of Experimental Psychology,* 1950, **40,** 494–503.

BUSS, A. H., BRADEN, W., ORGEL, A., & BUSS, E. H. Acquisition and extinction with different verbal reinforcement combinations. *Journal of Experimental Psychology,* 1956, **52,** 288–295.

BUSS, A. H., & BUSS, E. H. The effect of verbal reinforcement combinations on conceptual learning. *Journal of Experimental Psychology,* 1956, **52,** 283–287.

CAHILL, H. E., & HOVLAND, C. I. The role of memory in the acquisition of concepts. *Journal of Experimental Psychology,* 1960, **59,** 137–144.

CALLERTINE, M., & WARREN, J. M. Learning sets in human concept formation. *Psychological Reports,* 1955, **1,** 363–367.

CANTOR, J. H. Amount of pretraining as a factor in stimulus predifferentiation and performance set. *Journal of Experimental Psychology,* 1955, **50,** 180–184.

CAREY, J. E., & GOSS, A. E. The role of mediating verbal responses in the conceptual sorting behavior of children. *Journal of Genetic Psychology,* 1957, **90,** 69–74.

CARPENTER, F. The effect of differential reinforcement on concept formation. Unpublished doctoral dissertation, Indiana University, 1953.

CAUL, W. F., & LUDVIGSON, H. W. The effect of overlearning on response shifting. *Psychonomic Science,* 1964, **1,** 159–160.

COLBY, M. G., & ROBERTSON, J. B. Genetic studies in abstraction. *Journal of Comparative Psychology,* 1942, **33,** 385–401.

COLEMAN, E. B. Verbal concept learning as a function of instructions and dominance level. *Journal of Experimental Psychology,* 1964, **68,** 213–214.

CONANT, M. B. Utilization of negative instances in disjunctive concept learning. *Proceedings of the 74th Annual Convention of the American Psychological Association,* 1966, 37–38.

CONANT, M. B., & TRABASSO, T. Conjunctive and disjunctive concept formation under equal-information conditions. *Journal of Experimental Psychology,* 1964, **67,** 250–255.

CORAH, N. L. Color and form in children's perceptual behavior. *Perceptual and Motor Skills,* 1964, **18,** 313–316.

CROUSE, J. H., & DUNCAN, C. P. Verbal concept sorting as a function of response dominance and sorting method. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2,** 480–484.

DATTMAN, P. E., & ISRAEL, H. E. The order of dominance among conceptual capacities: An experimental test of Heidbreder's hypothesis. *Journal of Psychology,* 1951, **31,** 147–160.

DAVIDSON, R. S. The effects of symbols, shift, and manipulation upon the number of concepts attained. *Journal of Experimental Psychology,* 1952, **44,** 70–79.

DELLA-PIANA, G. M. Searching orientation and concept learning. *Journal of Educational Psychology,* 1957, **48,** 245–253.

DENNY, J. P. Effects of anxiety and intelligence on concept formation. *Journal of Experimental Psychology,* 1966, **72,** 596–602.

DENO, S. L. Multiple discrimination learning as a function of conceptual similarity among picture and word stimuli. Unpublished doctoral dissertation, University of Minnesota, 1965.

DETAMBEL, M. H., & STOLUROW, L. M. Stimulus sequence and concept learning. *Journal of Experimental Psychology,* 1956, **51,** 34–40.

DETERLINE, W. A. Verbal responses and concept formation. *Psychological Reports,* 1957, **3,** 372.

DICK, R. D. Formation and generalization of concepts as a function of similarity and amount of training. *American Journal of Psychology,* 1963, **76,** 607–615.

Dietz, D. The facilitating effects of words on discrimination and generalization. *Journal of Experimental Psychology,* 1955, **50,** 255–260.

Duncan, C. P. Mediation in verbal concept learning. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4,** 1–6.

Erlebacher, A., & Archer, E. J. Perservation as a function of degree of learning and percentage of reinforcement in card sorting. *Journal of Experimental Psychology,* 1961, **62,** 510–517.

Fenn, J. D., & Goss, A. E. The role of mediating verbal responses in conceptual sorting behavior of normals and paranoid schizophrenics. *Journal of Genetic Psychology,* 1957, **90,** 59–67.

Forgus, R. H., & Fowler, H. The order of dominance in concept attainment as affected by experience. *Journal of Psychology,* 1957, **44,** 105–108.

Freedman, J. L., & Mednick, S. A. Ease of attainment of concepts as a function of response dominance variance. *Journal of Experimental Psychology,* 1958, **55,** 463–466.

Freibergs, V., & Tulving, E. The effect of practice on utilization of information from positive and negative instances in concept identification. *Canadian Journal of Psychology,* 1961, **15,** 101–106.

Friedman, S. R. Developmental level and concept learning: Confirmation of an inverse relationship. *Psychonomic Science,* 1965, **2,** 34.

Fryatt, M. J., & Tulving, E. Interproblem transfer in identification of concepts involving positive and negative instances. *Canadian Journal of Psychology,* 1963, **17,** 106–117.

Furth, H. G. The influence of language on the development of concept formation in deaf children. *Journal of Abnormal and Social Psychology,* 1961, **63,** 386–389.

Gelfand, S. Effects of prior associations and task complexity upon the identification of concepts. *Psychological Reports,* 1958, **4,** 567–574.

Glanzer, M., Huttenlocher, J., & Clark, W. H. Systematic operations in solving concept problems: A parametric study of a class of problems. *Psychological Monographs,* 1963, **77**(1, Whole No. 564).

Goss, A. E., & Moylan, M. C. Conceptual blocksorting as a function of type and degree of mastery of discriminative verbal responses. *Journal of Genetic Psychology,* 1958, **93,** 191–198.

Grant, D. A. Perceptual versus analytical responses to the number of concept of a Weigl-type and sorting test. *Journal of Experimental Psychology,* 1951, **41,** 23–29.

Grant, D. A., & Berg, E. A. A behavioral analysis of degree of reinforcement and use of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology,* 1948, **38,** 404–411.

Grant, D. A., & Cost, J. R. Continuities and discontinuities in conceptual behavior in a card sorting problem. *Journal of General Psychology,* 1954, **50,** 237–244.

Grant, D. A., & Curran, J. F. Relative difficulty of number, form, and color concepts of a Weigl-type problem using unsystematic number cards. *Journal of Experimental Psychology,* 1952, **43,** 408–413.

Grant, D. A., Jones, O. R., & Tallantis, B. The relative difficulty of the number, form, and color concepts of a Weigl-type problem. *Journal of Experimental Psychology,* 1949, **39,** 552–557.

Green, E. J. Concept formation: A problem in human operant conditioning. *Journal of Experimental Psychology,* 1955, **49,** 175–180.

Griffin, M., & Beier, E. G. "Subliminal" prior solution cues in problem solving. *Journal of General Psychology,* 1961, **65,** 219–227.

Griffith, B. C., Spitz, H. H., & Lipman, R. S. Verbal mediation and concept formation in retarded and normal subjects. *Journal of Experimental Psychology,* 1959, **58,** 247–251.

Haygood, D. H. Audio-visual concept formation. *Journal of Educational Psychology,* 1965, **56,** 126–132.

Haygood, R. C., & Bourne, L. E., Jr. Forms of relevant stimulus redundancy in concept identification. *Journal of Experimental Psychology,* 1964, **67,** 392–397.

Haygood, R. C., & Devine, J. V. Effects of composition of positive category on concept learning. *Journal of Experimental Psychology,* 1967, **74,** 230–235.

Haygood, R. C., & Stevenson, M. Effects of number of irrelevant dimensions in nonconjunctive concept learning. *Journal of Experimental Psychology,* 1967, **74,** 302–304. (a)

Haygood, R. C., & Stevenson, M. Effects of proportion of positive instances upon concept learning. *Psychological Reports,* 1967, **20,** 179–182. (b)

Heidbreder, E. The attainment of concepts: I. Terminology and methodology. *Journal of General Psychology,* 1946, **35,** 173–189. (a)

Heidbreder, E. The attainment of concepts: II. The problem. *Journal of General Psychology,* 1946, **35,** 191–223. (b)

Heidbreder, E. The attainment of concepts: III. The process. *Journal of Psychology,* 1947, **24,** 93–138.

Heidbreder, E. The attainment of concepts: VI. Exploratory experiments on conceptualization at perceptual levels. *Journal of Psychology,* 1948, **26,** 193–216.

Heidbreder, E. The attainment of concepts: VII. Conceptual achievements during card sorting. *Journal of Psychology,* 1949, **27,** 3–39. (a)

Heidbreder, E. The attainment of concepts: VIII. The conceptualization of verbally indicated instances. *Journal of Psychology,* 1949, **27,** 263–309. (b)

Heidbreder, E., Bensley, M. L., & Ivy, M. The attainment of concepts: IV. Regularities and levels. *Journal of Psychology,* 1948, **25,** 299–329.

Heidbreder, E., & Overstreet, P. The attainment of concepts: V. Critical features and contexts. *Journal of Psychology,* 1948, **26,** 45–69.

Heidbreder, E., & Zimmerman, C. The attainment of concepts: IX. Sematic efficiency and concept attainment. *Journal of Psychology,* 1955, **40,** 325–335.

Holbrook, J. D. The influence of anxiety and task complexity on the learning of concepts. Unpublished doctoral dissertation, University of Nebraska, 1954.

Hovland, C. I., & Weiss, W. Transmission of information concerning concepts through positive and negative instances. *Journal of Experimental Psychology,* 1953, **45,** 175–182.

Hunt, E. B. Memory effects in concept learning. *Journal of Experimental Psychology,* 1961, **62,** 598–604.

Hunt, E. B., & Hovland, C. I. Order of consideration of different types of concepts. *Journal of Experimental Psychology,* 1960, **59,** 220–225.

Huttenlocher, J. Effects of manipulation of attributes on efficiency of concept formation. *Psychological Reports,* 1962, **10,** 503–509. (a)

Huttenlocher, J. Some effects of negative instances on the formation of simple concepts. *Psychological Reports,* 1962, **11,** 35–42. (b)

Jeffrey, W. E., & Kluppel, D. D. Mediational variables in concept formation. *Psychological Reports,* 1962, **10,** 191–202.

Johannsen, W. J. Concept identification under misinformative and subsequent informative feedback conditions. *Journal of Experimental Psychology,* 1962, **64,** 631–635.

Johnson, D. M., & O'Reilly, C. A. Concept attainment in children: Classifying and defining. *Journal of Educational Psychology,* 1964, **55,** 71–74.

Johnson, P. J. Factors affecting transfer in concept identification problems. *Journal of Experimental Psychology,* 1966, **72,** 655–660.

Juzak, T. The effects of praise and reproof on the generalization of learned concepts. *Journal of Psychology,* 1955, **39,** 329–340.

Kagan, J., & Lemkin, J. Form, color and size in children's conceptual behavior. *Child Development,* 1961, **32,** 25–28.

Kates, S. L., & Yudin, L. Concept attainment and memory. *Journal of Educational Psychology,* 1964, **55,** 103–109.

Katz, P. A. Role of irrelevant cues in the formation of concepts by lower-class children. *Journal of Educational Psychology,* 1968, **59,** 233–238.

Keele, S. W., & Archer, E. J. A comparison of two types of information in concept identification. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6,** 185–192.

Kendler, H. H., & Karasik, A. D. Concept formation as a function of competition between response produced cues. *Journal of Experimental Psychology,* 1959, **55,** 278–283.

Kendler, H. H., & Vineberg, R. The acquisition of compound concepts as a function of previous training. *Journal of Experimental Psychology,* 1954, **48,** 252–258.

Kendler, T. S. Concept formation. *Annual Review of Psychology,* 1961, **12,** 447–472.

Kendler, T. S., & Kendler, H. H. Effect of verbalization on reversal shifts in children. *Science,* 1961, **134,** 1619–1620.

Kendler, T. S., Kendler, H. H., & Wells, D. Reversal and nonreversal shifts in nursery school children. *Journal of Comparative and Physiological Psychology,* 1960, **53,** 83–88.

Kepros, P. G., & Bourne, L. E., Jr. Identification of biconditional concepts: Effects of number of relevant and irrelevant dimensions. *Canadian Journal of Psychology,* 1966, **20,** 198–207.

Kirloskar, S., & Parameswaran, E. G. Effect of irrelevant material on concept formation. *Research Bulletin of the Department of Psychology,* Osmania University, 1967, **3,** 18–25.

Klausmeier, H. J., & Meinke, D. L. Concept attainment as a function of instructions concerning the stimulus material, a strategy and a principle for securing information. *Journal of Educational Psychology,* 1968, **59,** 215–222.

Klugh, E., & Roehl, K. Developmental level and concept learning: Interaction of age and complexity. *Psychonomic Science,* 1965, **2,** 385–386.

Kluppel, D. D., & Jeffrey, W. E. Stimulus labelability as a factor in concept formation. *Psychological Reports,* 1964, **15,** 918.

Kofsky, E. The effect of verbal training on concept identification in disadvantaged children. *Psychonomic Science,* 1967, **7,** 365–366.

Komachiya, M. A note on concept attainment. *Journal of Psychology,* 1957, **43,** 261–263.

Krossner, W. J. Density and dimensionality in the attainment of conjunctive concepts. *Psychonomic Science,* 1966, **6,** 265–266.

Kurtz, K. H., & Hovland, C. I. Concept learning with differing sequences of instances. *Journal of Experimental Psychology,* 1956, **51,** 239–243.

Laughlin, P. R. Selection strategies in concept attainment as a function of number of persons and stimulus display. *Journal of Experimental Psychology,* 1965, **70,** 323–327.

Laughlin, P. R. Selection strategies in concept attainment as a function of number of relevant problem attributes. *Journal of Experimental Psychology,* 1966, **71,** 773–776.

Laughlin, P. R., & Doherty, M. A. Discussion versus memory in cooperative group concept attainment. *Journal of Educational Psychology,* 1967, **58,** 123–128.

Laughlin, P. R., & Jordan, R. M. Selection strategies in conjunctive, disjunctive, and biconditional concept attainment. *Journal of Experimental Psychology,* 1967, **75,** 188–193.

Lee, L. C. Concept utilization in pre-school children. *Child Development,* 1965, **36,** 221–227.

Long, B. H. Predecisional search in concept formation: The effects of problem complexity. *Psychological Record,* 1965, **15,** 197–202.

Long, L. Conceptual relationships in children: The concept of roundness. *Journal of Genetic Psychology,* 1940, **57,** 289–315.

Lordahl, D. S. Concept identification using simultaneous auditory and visual signals. *Journal of Experimental Psychology,* 1961, **62,** 283–290.

Mayzner, M. S. Verbal concept attainment: A function of the number of positive and negative instances presented. *Journal of Experimental Psychology,* 1962, **63,** 314–319.

MAYZNER, M. S., & TRESSELT, M. E. Verbal concept attainment: A function of the number of strength of positive instances presented. *Journal of Psychology,* 1962, **53,** 469–474.

MCCREARY, J. Y. B. Factors affecting the attainment of conjunctive and disjunctive concepts in third grade children. Unpublished doctoral dissertation, University of Pittsburgh, 1963.

MCGUIGAN, F. J. Incidental learning in the formation of concepts. *American Journal of Psychology,* 1958, **71,** 539–547.

MEDNICK, S. A., & FREEDMAN, J. L. Facilitation of concept formation through mediated generalization. *Journal of Experimental Psychology,* 1960, **60,** 278–283.

MEDNICK, S. A., & HALPERN, S. Ease of concept attainment as a function of associative rank. *Journal of Experimental Psychology,* 1962, **64,** 628–630.

MEYER, W. J., & OFFENBACH, S. I. Effectiveness of reward and punishment as a function of task complexity. *Journal of Comparative and Physiological Psychology,* 1962, **55,** 532–534.

MEYER, W. J., & SEIDMAN, S. B. Age differences in the effectiveness of different reinforcement combinations on the acquisition and extinction of a simple concept learning problem. *Child Development,* 1960, **31,** 419–424.

MEYER, W. J., & SEIDMAN, S. B. Relative effectiveness of different reinforcement combinations on concept learning of children at two developmental levels. *Child Development,* 1961, **32,** 117–127.

MILGRAM, N. A., & NOCE, J. S. Relevant and irrelevant verbalization in discrimination and reversal learning by normal and retarded children. *Journal of Educational Psychology,* 1968, **59,** 169–175.

MORIN, R. E. Factors influencing rate and extent of learning in the presence of misinformative feedback. *Journal of Experimental Psychology,* 1955, **49,** 343–351.

MORRISETT, L., JR., & HOVLAND, C. I. A comparison of three varieties of training in human problem solving. *Journal of Experimental Psychology,* 1959, **58,** 52–55.

NADELMAN, L. Influence of concreteness and accessibility on concept thinking. *Psychological Reports,* 1957, **3,** 189–212.

NAMIKAS, G. Concept identification as a function of relevance of pretraining and percentage of informative feedback. *Psychonomic Science,* 1967, **8,** 261–262.

NEISSER, U., & WEENE, P. Hierarchies in concept attainment. *Journal of Experimental Psychology,* 1962, **64,** 640–645.

NEWMAN, S. E. Effects of contiguity and similarity on the learning of concepts. *Journal of Experimental Psychology,* 1956, **52,** 349–353.

NICHOLSON, E. Concept attainment of male high school freshmen. Unpublished doctoral dissertation, University of Pennsylvania, 1966.

NORCROSS, K. J., & SPIKER, C. C. The effects of type of stimulus pretraining and discrimination performance in preschool children. *Child Development,* 1957, **28,** 79–84.

O'CONNEL, D. C. Concept learning and verbal control under partial reinforcement and subsequent reversal or nonreversal shifts. *Journal of Experimental Psychology,* 1965, **69,** 144–151.

ODOM, R. V. Concept identification and utilization among children of different ages. *Journal of Experimental Child Psychology,* 1966, **4,** 309–316.

OLSON, L. A. Concept attainment of high school sophomores. *Journal of Educational Psychology,* 1963, **54,** 213–216.

OSEAS, L., & UNDERWOOD, B. J. Studies in distributed practice: V. Learning and retention of concepts. *Journal of Experimental Psychology,* 1952, **43,** 143–148.

OSLER, S. F., & FIVEL, M. W. Concept attainment: I. The role of age and intelligence in concept attainment by induction. *Journal of Experimental Psychology,* 1961, **62,** 1–8.

OSLER, S. F., & KOFSKY, E. Stimulus uncertainty as a variable in the development of conceptual ability. *Journal of Experimental Child Psychology,* 1965, **2,** 264–279.

OSLER, S. F., & KOFSKY, E. Structure and strategy in concept learning. *Journal of Experimental Psychology,* 1966, **4,** 198–209.

OSLER, S. F., & SHAPIRO, S. L. Studies in concept attainment: IV. The role of partial reinforcement as a function of age and intelligence. *Child Development,* 1964, **35,** 623–633.

OSLER, S. F., & TRAUTMAN, G. E. Concept attainment: II. Effects of stimulus complexity upon concept attainment at two levels of intelligence. *Journal of Experimental Psychology,* 1961, **62,** 9–13.

OSLER, S. F., & WEISS, S. R. Studies in concept attainment: III. Effect of instructions at two levels of intelligence. *Journal of Experimental Psychology,* 1962, **63,** 528–533.

OTTO, W. The differential effects of verbal and pictorial representations of stimuli upon responses evoked. *Journal of Verbal Learning and Verbal Behaviors,* 1962, **1,** 192–196.

OTTO, W. Hierarchial responses elicited by verbal and pictorial stimuli. *American Educational Research Journal,* 1964, **1,** 241–248.

PENNEY, R. K., & LUPTON, A. A. Children's discrimination learning as a function of reward and punishment. *Journal of Comparative and Physiological Psychology,* 1961, **54,** 449–451.

PETERSON, M. J. Some effects of the percentage of relevant cues and presentation methods on concept identification. *Journal of Experimental Psychology,* 1962, **64,** 623–627.

PIAGET, J. *The child's conception of physical causality.* New York: Harcourt Brace, 1930.

PISHKIN, V. Effects of probability of misinformation and number of irrelevant dimensions upon concept identification. *Journal of Experimental Psychology,* 1960, **59,** 371–378.

Pishkin, V. Stimulus and response tendencies in concept identification. *Perceptual and Motor Skills,* 1961, **13,** 295–304. (a)

Pishkin, V. Transmission of information as a function of misinformation feedback distribution. *Psychological Reports,* 1961, **9,** 255–263. (b)

Pishkin, V. Dimension availability with antecedent success or failure in concept identification. *Psychonomic Science,* 1965, **2,** 69–70.

Pishkin, V. Availability of feedback-corrected error instances in concept learning. *Journal of Experimental Psychology,* 1967, **73,** 318–319.

Pishkin, V., & Blanchard, R. J. Stimulus and social cues in concept identification of schizophrenics and normals. *Journal of Abnormal and Social Psychology,* 1963, **67,** 454–463.

Pishkin, V., & Blanchard, R. J. Auditory concept identification as a function of subjects' sex and stimulus dimension. *Psychonomic Science,* 1964, **1,** 177–178.

Pishkin, V., & Shurley, J. T. Auditory dimensions and irrelevant information in concept identification of males and females. *Perceptual and Motor Skills,* 1965, **20,** 673–683.

Pishkin, V., & Wolfgang, A. Number and type of available instances in concept learning. *Journal of Experimental Psychology,* 1965, **69,** 5–8.

Pishkin, V., Wolfgang, A., & Rasmussen, E. Age, sex, amount, and type of memory information in concept learning. *Journal of Experimental Psychology,* 1967, **73,** 121–124.

Pollack, I. Speed of classification of words into superordinate categories. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2,** 159–165.

Rankin, R. E. Inter-concept stimulus similarity and formation of simple concepts. Unpublished doctoral dissertation, Ohio State University, 1954.

Rasmussen, E. A., & Archer, E. J. Concept identification as a function of language pretraining and task complexity. *Journal of Experimental Psychology,* 1961, **61,** 437–441.

Reed, H. B. Factors influencing the learning and retention of concepts. I. The influence of set. *Journal of Experimental Psychology,* 1946, **36,** 71–87. (a)

Reed, H. B. The learning and retention of concepts. II. The influence of length of series. III. The origin of concepts. *Journal of Experimental Psychology,* 1946, **36,** 166–179. (b)

Reed, H. B. The learning and retention of concepts. IV. The influence of the complexity of the stimuli. *Journal of Experimental Psychology,* 1946, **36,** 252–261. (c)

Reed, H. B. The learning and retention of concepts. V. The influence of form of presentation. *Journal of Experimental Psychology,* 1950, **40,** 504–511.

Restle, F., & Emmerich, D. Memory in concept attainment: Effects of giving several problems concurrently. *Journal of Experimental Psychology,* 1966, **71,** 794–799.

Richardson, J., & Bergum, B. O. Distributed practice and rote learning in concept formation. *Journal of Experimental Psychology,* 1954, **47,** 442–446.

Romanow, C. V. Anxiety level and ego involvement as factors in concept formation. *Journal of Experimental Psychology,* 1958, **56,** 166–173.

Runquist, W. N., & Hutt, V. H. Verbal concept learning in high school students with pictorical and verbal representation of stimuli. *Journal of Educational Psychology,* 1961, **52,** 108–111.

Russell, D. H., & Saadeh, I. Q. Qualitative levels in children's vocabularies. *Journal of Educational Psychology,* 1962, **53,** 170–174.

Saltz, E., & Sigel, I. E. Concept overdiscrimination in children. *Journal of Experimental Psychology,* 1967, **73,** 1–8.

Saunders, P. S., & Neimark, E. D. Effects of misinformative feedback and awareness of misinformation on concept attainment strategy. *Proceedings of the 75th Annual Convention of the American Psychological Association,* 1967, **2,** 49–50.

Sax, G. Concept acquisition as a function of differing schedules and delays of reinforcement. *Journal of Educational Psychology,* 1960, **51,** 32–36.

Scholl, P. A. Visual concept attainment as affected by stimulus complexity and selected communicator strategies. Unpublished EdD dissertation, Indiana University, 1966.

Schulz, R. W., Miller, R. L., & Radtke, R. C. The role of instance contiguity and dominance in concept attainment. *Journal of Verbal Learning and Verbal Behavior,* 1963, **1,** 432–435.

Schvaneveldt, R. W. Concept identification as a function of probability of positive instances and number of relevant dimensions. *Journal of Experimental Psychology,* 1966, **72,** 649–654.

Schwartz, B. D. Certain relations between verbalization and concept formation: The distinction of ideas by words. Unpublished doctoral dissertation, Princeton University, 1948.

Schwartz, S. H. Trial-by-trial analysis of processes in simple disjunctive concept-attainment tasks. *Journal of Experimental Psychology,* 1966, **72,** 456–465.

Sechrest, L., & Kaas, J. S. Concept difficulty as a function of stimulus similarity. *Journal of Educational Psychology,* 1965, **56,** 327–333.

Sechrest, L., & Wallace, J. Assimilation and utilization of information in concept attainment under varying conditions of information presentation. *Journal of Educational Psychology,* 1962 **53,** 157–164.

Semme, L. R. M., & Warren, J. M. Proactive facilitation of performance on the Wisconsin card sorting test. *Psychological Reports,* 1955, **2,** 433–436.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. Learning and memorization of classifications. *Psychological Monographs,* 1961, **75**(4, Whole No. 517).

Shepard, W. O., & Schaeffer, M. The effect of concept knowledge on discrimination learning. *Child Development,* 1956, **27,** 173–178.

Shore, E., & Sechrest, L. Concept attainment as a function of the number of positive instances presented. *Journal of Educational Psychology,* 1961, **52,** 303–307.

Sigel, I. E. Developmental trends in the abstraction ability of children. *Child Development,* 1953, **24,** 131–144.

Smoke, K. L. An objective study of concept formation. *Psychological Monographs,* 1932, **42**(4, Whole No. 191).

Smoke, K. L. Negative instances in concept learning. *Journal of Experimental Psychology,* 1933, **16,** 583–588.

Spiker, C. C. Stimulus pretraining and subsequent performance in the delayed reaction experiment. *Journal of Experimental Psychology,* 1956, **52,** 107–111.

Spiker, C. C. Verbal factors in the discrimination learning of children. *Society for Research in Child Development,* 1963, **28,** 53–69.

Stern, C. Labeling and variety in concept identification with young children. *Journal of Educational Psychology,* 1965, **56,** 235–240.

Suchman, R. G., & Trabasso, T. Color and form preference in young children. *Journal of Experimental Child Psychology,* 1966, **3,** 177–187. (a)

Suchman, R. G., & Trabasso, T. Stimulus preference and cue function in young children's concept attainment. *Journal of Experimental Child Psychology,* 1966, **3,** 188–198. (b)

Sullivan, P. W. The effects of verbal reward and verbal punishment on concept elicitation in children. Unpublished doctoral dissertation, Wayne State University, 1960.

Tavrow, W. N. The effects of practice and of previous familiarity with the universe in forming concepts from negative instances. Unpublished doctoral dissertation, Columbia University, 1965.

Thysell, R. V., & Schulz, R. W. Concept-utilization as a function of the strength of relevant and irrelevant associations. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3,** 203–208.

Trabasso, T. R. Stimulus emphasis and all-or-none learning in concept identification. *Journal of Experimental Psychology,* 1963, **65,** 398–406.

Trabasso, T. R. Verification of conceptual rules. *Proceedings of the 75th Annual Convention of the American Psychological Association,* 1967, **2,** 41–42.

Turnure, C., & Wallach, L. The influence of contextual variation on the differentiation of parts from wholes. *American Journal of Psychology,* 1962, **78,** 481–485.

Underwood, B. J. Studies of distributed practice: XV. Verbal concept learning as a function of intralist interference. *Journal of Experimental Psychology,* 1957, **54,** 33–40.

Underwood, B. J., & Richardson, J. Verbal concept learning as a function of instructions and dominance level. *Journal of Experimental Psychology,* 1956, **51,** 229–238.

Walker, C. M., & Bourne, L. E., Jr. The identification of concepts as a function of amounts of relevant and irrelevant information. *American Journal of Psychology,* 1961, **74,** 410–417.

Wallace, J., Jr. Concept dominance, type of feedback, and intensity of feedback as related to concept attainment. *Journal of Educational Psychology,* 1964, **55,** 159–166.

Wallace, J., Jr., & Sechrest, L. Relative difficulty of conjunctive and disjunctive concepts. *Journal of Psychological Studies,* 1961, **12,** 97–104.

Wallach, L. The complexity of concept attainment. *American Journal of Psychology,* 1962, **75,** 277–283.

Weiner, A. B. The effect of a learning strategy on memory, information, and intelligence in concept attainment. Unpublished doctoral dissertation, University of Massachusetts, 1967.

Weir, M. W., & Stevenson, H. W. Effect of verbalization in children's learning as a function of chronological age. *Child Development,* 1959, **30,** 173–178.

Wells, H. Stimulus compounding and the attainment of conjunctive and disjunctive concepts. *Perceptual and Motor Skills,* 1965, **21,** 767–770.

Wells, H., & Deffenbacher, K. Conjunctive and disjunctive concept learning in human and squirrel monkeys. *Canadian Journal of Psychology,* 1967, **21,** 301–308.

Wells, H., & Watson, D. Strategy training and practice in disjunctive concept attainment. *Psychological Reports,* 1965, **17,** 925–926.

Wenzel, B. M., & Flurry, C. The sequential order of concept attainment. *Journal of Experimental Psychology,* 1948, **38,** 547–557.

Wesley, E. L. Perservative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *Journal of Abnormal and Social Psychology,* 1953, **48,** 129–134.

Whitman, J. R., & Garner, W. R. Concept learning as a function of form of internal structure. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2,** 195–202.

Wittrock, M. C. Verbal stimuli in concept formation: Learning by discovery. *Journal of Educational Psychology,* 1963, **54,** 183–190.

Wolff, J. L. Concept attainment, intelligence, and stimulus complexity: An attempt to replicate Osler and Trautman (1961). *Journal of Experimental Psychology,* 1967, **73,** 488–490. (a)

Wolff, J. L. Effects of verbalization and pretraining on concept attainment by children in two mediation categories. *Journal of Educational Psychology,* 1967, **58,** 19–26. (b)

Wolfgang, A. Effects of social cues and task complexity in concept identification. *Journal of Educational Psychology,* 1967, **58,** 36–40.

Wolfgang, A., Pishkin, V., & Lundy, R. M. Anxiety and misinformation feedback in concept attainment. *Perceptual and Motor Skills,* 1962, **14,** 135–143.

Wylie, A. A. Concept formation in pre-school children. Unpublished doctoral dissertation, University of Houston, 1958.

Youniss, J., & Furth, H. G. Learning of logical connectives by adolescents with single and multiple instances. *Journal of Educational Psychology,* 1967, **58,** 222–230.

Yudin, L., & Kates, S. L. Concept attainment and adolescent development. *Journal of Educational Psychology,* 1963, **54,** 177–182.

(Received May 6, 1970)

UNIVERSITY OF VICTORIA

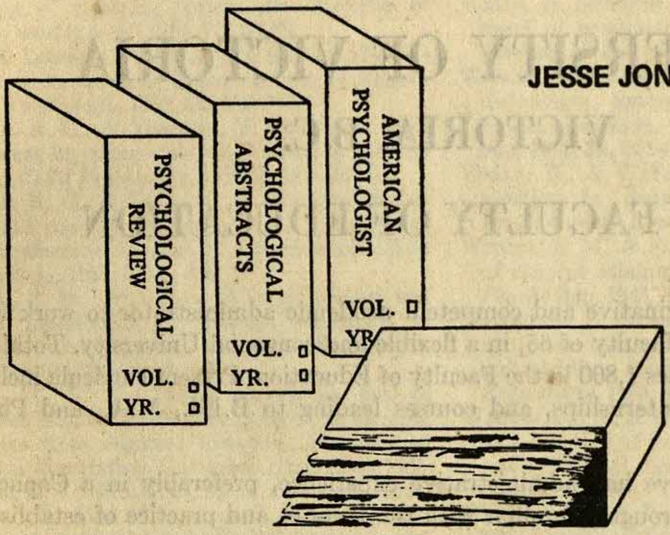
VICTORIA, B.C.

DEAN, FACULTY OF EDUCATION

Opportunity for imaginative and competent academic administrator to work with a strong, well-qualified faculty of 65, in a flexible and congenial University. Total enrollment of 6,000, includes 1,800 in the Faculty of Education. Present curricula include diploma programmes, internships, and courses leading to B.Ed., M.A., and Ph.D. degrees.

The Dean should have had administrative experience, preferably in a Canadian university, must be thoroughly familiar with the theories and practice of established and innovative methods of teacher training and of research in education, and must have demonstrated the capacity for educational and personal leadership.

Please reply in confidence to the President, furnishing curriculum vitae, names of five references, a statement of your career aspirations and outlook, and salary requirements.



JESSE JONES VOLUME FILES

Send orders to:
JESSE JONES CORP.
P.O. Box 5120
Philadelphia, Pa. 19141

Only $3.50 ea.
3 for $10.00
6 for $19.00 P.P.D.

Now, the journals of the American Psychological Association can become an attractive permanent part of your professional library. These famous Jesse Jones volume files, especially designed to keep your copies orderly, readily accessible for future references — guard against soiling, tearing, wear or misplacement of copies.

This durable file will support 150 lbs. Looks and feels like leather and is washable. The 23-carat gold lettering makes it a fit companion for the most costly binding.

JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 62 August 1971 Number 4

CONTENTS

Conditioning of Verbal Behavior as a Function of Age, Social Class, and Type of Reinforcement 279
JOHN D. CRADLER AND DWIGHT L. GOODWIN

Effectiveness of Student-to-Student Counseling of the Academic Adjustment to Potential College Dropouts 285
WILLIAM F. BROWN, NATHAN O. WEHE, VERNON G. ZUNKER, AND WARREN L. HASLAM

Effects of Structural Complexity of Bigrams on Paired-Associate Learning and Retention of Children 290
ROBERT A. HICKS AND JAMES PACKWOOD

Effects of Intelligence, Information Processing, and Mediation Conditions on Conceptual Learning 293
JIM MILLHAM, LEONARD I. JACOBSON, AND STEPHEN E. BERGER

Differential Response to Inductive and Deductive Instructional Procedures 300
MARY LOU KORAN

Effects of Encounter with Verbal Units in Prose Contexts on Subsequent Associative Learning 308
ROGER H. BRUNING AND LARRY J. LANTINGA

Thematic Prompting in Paired-Associate Learning 315
RICHARD C. ANDERSON, JAMES M. ROYER, RAYMOND W. KULHAVY, SYLVIA D. THORNBURG, AND LAURA L. KLEMT

Retention and Transfer of Mediation Set in Paired-Associate Learning of Educable Retarded Children 322
DOROTHEA ROSS

Adjustment of Anglo-American and Mexican-American Pupils in Self-Contained and Team-Teaching Classrooms 328
LINDA SCHMIDT AND JUNE GALLESSICH

Effect of Availability on the Relationship of Memory Abilities to Performance in Multiple-Category Concept Tasks 333
DANIEL D. BLAINE AND JACK L. DUNHAM

Effects of Verbal and Pictorial Elaborations on Associative Learning and Response Learning in a Children's Paired-Associate Task 339
STEVE LYNCH AND WILLIAM D. ROHWER, JR.

Relationships of Teaching Method, Socioeconomic Status, and Intelligence in Concept Formation 345
JOHN T. GUTHRIE

Effect of Irrelevant Attributes and Irrelevant Operations on Rule Learning 352
JOSEPH M. SCANDURA AND DONALD J. VOORHIES

Interaction among Anxiety, Stress, Response Mode, and Familiarity of Subject Matter on Achievement from Programmed Instruction 357
SIGMUND TOBIAS AND THEODORE ABRAMSON

© 1971 by the American Psychological Association, Inc.

Journal of Educational Psychology
1971, Vol. 62, No. 4, 279–284

CONDITIONING OF VERBAL BEHAVIOR AS A FUNCTION OF AGE, SOCIAL CLASS, AND TYPE OF REINFORCEMENT[1]

JOHN D. CRADLER AND DWIGHT L. GOODWIN[2]
South San Francisco Unified School District
San Jose State College

Employing a 2 × 2 × 3 factorial design, 144 children participated in an investigation of the independent and interactive effects of three types of reinforcement, two age levels, and two social classes on the conditioning of the use of one of four pronouns in the construction of sentences. Awareness of the reinforcement contingency was determined by an interview. A three-way interaction between age level, social class, and type of reinforcement was not significant. However, two significant two-way interactions were obtained. In addition, awareness of reinforcement contingencies significantly facilitated conditioning.

Previous research has suggested that there are significant differences in the relative effectiveness of material and nonmaterial incentives for children of different social class backgrounds. For example, Terrell, Durkin, and Weisley (1959) found that lower-class subjects learned a size and shape discrimination task faster when given a material reward, while middle-class subjects learned to make the same discrimination faster when rewarded by a light flash, a highly symbolic reinforcer. Similar findings were obtained by Zigler and DeLabry (1962), Staats and Butterfield (1955), Risley (1968), and Cameron and Storm (1965) in that these authors found lower-class subjects to be more responsive to material than to social or abstract, symbolic reinforcement. The susceptibility of lower-class children to material reinforcers has been interpreted as resulting from a lack of social reinforcement dispensed by parents for academic learning (Katz, 1967) and deprivation of tangible (material) rewards (Douvan, 1956).

In summary, it is evident that lower-class children tend to prefer material over social and/or symbolic reinforcement. However, studies have shown that when lower-class children were limited to choices between social or symbolic reinforcers, they preferred social rewards. For example, Rosehan (1966), Zigler and Kanzer (1962), and Herold (1967) found that lower-class subjects conditioned to a significantly greater degree with social approval than with knowledge of a correct response, while the reverse was found for middle-class subjects. Zigler and Kanzer (1962) explained the lower-class subjects' increased responsiveness to social approval as a result of a lag in the socialization process for these subjects. That is, unlike middle-class children, lower-class children have not developed internalized self-reinforcing behavior in that the environment has typically not paired external social and/or material reinforcement with such intangible cues as grades, tokens, or knowledge of correct responding. As a result, lower-class children may remain somewhat dependent on immediate primary reinforcement, while middle-class children have learned to reward themselves internally when abstract external cues are present.

There have been several attempts to verify the suggestion of Zigler and Kanzer (1962) that as children mature, they become more responsive to less concrete and more symbolic rewards. Lewis, Wall, and Aronfreed (1963) and Rosenhan and Greenwald (1965) found that younger children were more responsive to social (praise) reinforcers, while older children were more responsive to symbolic (light-flash) reinforcement. Somewhat similar findings have been

---

[1] This article is based on a thesis submitted by the first author, under the second author's supervision, to San Jose State College in partial fulfillment of the requirements for the master's degree.

[2] Requests for reprints should be sent to Dwight L. Goodwin, Department of Psychology, San Jose State College, 125 South Seventh Street, San Jose, California 95114.

Copyright © 1971 by the American Psychological Association, Inc.

reported by Witryol and Ormsby (1961), Sullivan (1960), Jeffrey and Skager (1962), and Walters and Foote (1962). Most of these authors concluded that as children progress from childhood to adolescence, they become increasingly less responsive to social reinforcement dispensed by adults, and become more responsive to less concrete, more abstract symbolic reinforcement.

Most of the previously mentioned research on social class and incentive valuations has dealt only with young children at a given age level, and has shown that lower-class subjects from approximately preschool to third grade tend to place a higher value on social than material or symbolic reinforcement. In contrast, the research by Lewis et al. (1963) and Rosenhan et al. (1965) indicates that as middle-class children reach adolescence, they place more value on symbolic than social reinforcers. If young lower-class children are more responsive to material or concrete types of reinforcement than middle-class subjects because they are developmentally behind middle-class subjects (Zigler & Kanzer, 1962), it would seem that as they mature, they would become more responsive to social and/or symbolic reinforcers. The latter assumption received support from an earlier study by Cradler (1968) which showed that sixth-grade lower-class subjects were more responsive to social reinforcement than their second-grade counterparts.

In general, the literature suggests that as children increase in age, they undergo a transition from being responsive to concrete or social reinforcers to more abstract or symbolic incentives. Also, the research supports the idea that lower-class children are developmentally behind middle-class children with regard to incentive preferences in that they tend to prefer material or social reward to the more abstract incentives preferred by middle-class children. The present study was designed to investigate simultaneously the effects of age level, social class, and type of reinforcement on the conditioning of verbal behavior. It was predicted that second-grade lower-class subjects would be more responsive to material than to social or symbolic reinforcement, while middle-class second graders were expected to be more responsive to social than to material or symbolic reinforcement. It was also predicted that at Grade 6, all subjects would place higher value on less concrete reinforcers, with lower-class subjects being more responsive to social than to material or symbolic reinforcement, and with middle-class subjects being more responsive to symbolic rewards than to either social or material types of reinforcement.

METHOD

Subjects

The subjects consisted of 72 second-grade (43 males and 29 females) and 72 sixth-grade (35 males and 37 females) children selected from Christopher and Miner Schools in the Oak Grove Elementary School District in San Jose, California. Half of the subjects at each grade level were middle-class, and the other half were lower-class. To be identified as of lower socioeconomic status, a child had to come from a home where the occupation of the wage earner fell at or below Classes IV or V on the Hollingshead and Redlich Index of Social Position (Hollingshead & Redlich, 1958). The middle-class children were selected from homes where the wage earner fell within Class II or III. The experimenters consisted of two Caucasian males, ages 26 and 27.

Materials

The experimental materials included 100 3 × 5 inch cards with a different verb of the simple past tense typed in the center of each card. On a separate card, typed in a center row, there appeared four personal pronouns: WE, SHE, HE, THEY. The 100 verbs were commonly used words, arbitrarily chosen from second-grade readers. Tally sheets, divided into five columns, were used for tallying responses for the five blocks of 20 sentences to be composed by the subjects. Several packages of plain chocolate M&M candies were included to serve as the material reinforcement.

Procedure

The experimental design allows for comparing the effects of simultaneously varying three dichotomous variables: treatment (material versus praise versus symbolic reinforcement), social class (lower versus middle), and grade level (Grade 2 versus Grade 6). With independent replications for each grade level, all subjects of each social class were assigned randomly and in equal numbers to each of the three treatment conditions. As indicated by the format of a 2 × 2 × 3 factorial design, there was a total of twelve separate cells available for assignment of 12 subjects to each cell.

The verbal conditioning method described by

Taffel (1955) was employed in this study. The task requires each subject to compose sentences using each of 100 past tense verbs in conjunction with any one of four pronouns. The experimenter first placed a card with the four pronouns in front of the subject. He then read the following instructions to the subject:

> The purpose of this experiment is to study how people make up sentences. I will show you a number of cards each of which contains a different word typed in the center. You are to make up any sentence which you can think of, using the word on this card [the experimenter points to the card with the four pronouns]. It doesn't matter whether your sentence is long or short, true or false, simple or complicated, but it is important that you answer with the first sentence which enters your mind. At first, you may find this difficult, but as we continue, it will become easier, so don't feel discouraged.

Next, the cards with the verbs were placed in front of the subject one at a time. For the first 20 cards or trials, the experimenter said nothing. This was Block 1, the operant block, determining the base rate of response. (This method of obtaining an operant level was used by Baer & Goldfarb, 1962). By not using reinforcement during Block 1, each subject was his own control. Blocks 2, 3, 4, and 5 consisted of the next 80 trials, 20 per block, and were the acquisition blocks. For these 80 trials, each time a subject composed a sentence using "they," the experimenter rewarded him. The subjects in the material reinforcement groups were rewarded by being given an M&M candy by the experimenter immediately after using "they" in a sentence. The subjects in the praise reinforcement groups were rewarded by the experimenter saying "good." The subjects in the symbolic reinforcement groups simply earned a plus mark, placed by the experimenter on a blank sheet of paper, which could be easily seen by the subject, immediately after using "they" in a sentence. Only positive reinforcement on a continuous schedule was used. The pronoun "they" was chosen to be reinforced because it has a relatively low base rate, which provides the opportunity for increase with conditioning. Finally, whenever a subject was unable to read a word presented to him, the experimenter pronounced the word.

After completion of the task, each subject was questioned in order to assess his degree of awareness of what was taking place during the experiment. He was asked the following questions: (*a*) What do you think I was doing? (*b*) Why do you think I asked you to make up sentences? (*c*) Did you notice I said good (gave you candy or plus marks)? The experimenter recorded all responses on a prepared form, hidden from the subjects' view.

RESULTS

The data used for the statistical analysis were obtained by subtracting Block 1 (the base rate) from Block 5 for each subject. Thus, a difference score was obtained for each subject as an index of the degree of conditioning. This method of obtaining conditioning scores was used by Spielberger, Levin, and Shepard (1962). Since there were some negative difference scores, the lowest negative score being −4, a constant of 4 was added to each of the scores in order to transform all scores to positive numbers for expediency of computation. The means and standard deviations of the transformed difference scores for each of the 12 groups or cells are presented in Table 1. The mean difference scores for all 12 groups are illustrated in Figure 1.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF DIFFERENCE SCORES BY SOCIAL CLASS, GRADE LEVEL, AND TYPE OF REINFORCEMENT

Social class	Grade level					
	Grade 2			Grade 6		
	Material	Praise	Symbolic	Material	Praise	Symbolic
Middle						
M	5.66	6.64	6.58	5.75	8.25	9.66
SD	2.32	3.28	2.98	1.63	4.12	3.37
Lower						
M	8.41	5.41	5.25	6.75	7.00	6.33
SD	4.23	3.63	2.74	4.20	4.16	3.22

A three-way analysis of variance was computed in order to compare the simultaneous effects of social class, grade level, and

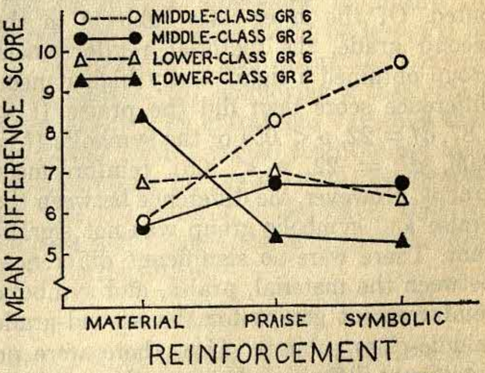


FIG. 1. Type of reinforcement. (Mean difference scores of "they" responses to material, praise, and symbolic reinforcement for lower-class and middle-class subjects at Grades 2 and 6.)

TABLE 2

ANALYSIS OF VARIANCE FOR SOCIAL CLASS TYPE OF REINFORCEMENT AND GRADE LEVEL

Source	*df*	*MS*	*F*
Social class (A)	1	12.250	.957
Grade level (B)	1	25.000	1.954
Type of reinforcement (C)	2	1.465	.114
A × B	1	11.111	.868
A × C	2	41.646	3.255*
B × C	2	47.396	3.704*
A × B × C	2	5.298	.414
Error	132	12.792	

* $p < .05$.

type of reinforcement. The results of this analysis are presented in Table 2. There were no significant main effects of social class ($F = .957$, $df = 1/132$, $p > .05$), grade level ($F = 1.95$, $df = 1/132$, $p > .05$), or type of reinforcement ($F = .114$, $df = 1/132$, $p > .05$). The interaction between social class and grade level was not significant ($F = .868$, $df = 1/132$, $p > .05$). However, there was a significant interaction between social class and type of reinforcement ($F = 3.255$, $df = 2/132$, $p > .05$), and between grade level and type of reinforcement ($F = 3.704$, $df = 2/132$, $p > .05$). The three-way interaction between social class, grade level, and type of reinforcement was not significant ($F = .414$; $df = 2/132$, $p > .05$). In order to determine whether or not there were any significant differences between individual cells and to test specific aspects of the hypotheses, *t* tests were computed. Of the lower-class subjects in the second grade, the material reinforcement group obtained a significantly higher mean difference score than did the praise ($t = 1.87$, $df = 22$, $p < .05$) or the symbolic ($t = 2.09$, $df = 22$, $p < .05$) reinforcement groups. However, the difference between the praise and symbolic group was not significant. There were no significant differences between the material, praise, and symbolic reinforcement groups for the second-grade middle-class subjects. Also, there were no significant differences between the material, praise, and symbolic reinforcement groups for the sixth-grade lower-class subjects. Of the sixth-grade middle-class subjects, the material reinforcement group obtained a significantly lower mean difference than did the praise ($t = 1.92$, $df = 22$, $p < .05$). The only significant difference between grade level for middle-class subjects was a significantly higher mean for the symbolic reward group at Grade 6 than at Grade 2 ($t = 2.26$, $df = 22$, $p < .05$). There were no significant grade level differences in mean difference scores for lower-class subjects. The only significant social class difference for Grade 2 was a significantly higher mean for the lower-class material reward group than for the middle-class material reward ($t = 2.01$, $df = 22$, $p < .05$). At Grade 6, the middle-class symbolic reward group obtained a significantly higher mean than did the lower-class symbolic reward group ($t = 2.37$, $df = 22$, $p < .05$).

Forty-three of the subjects were aware of the reinforcement contingencies. At Grade 2, 11 middle-class and 5 lower-class subjects were aware, and at Grade 6, 16 middle-class and 11 lower-class subjects were aware of the reinforcement contingencies. There were indications that the degree of conditioning was related to the subjects' awareness of the reinforcement contingencies. For this reason, a *t* test was computed in order to determine whether or not the mean of the difference scores for the aware subjects was significantly greater than the mean of the difference scores of the subjects who were not aware. A highly significant difference in mean difference scores between the aware and nonaware subjects was obtained ($t = 4.28$, $df = 121$, $p < .005$).

DISCUSSION

The present study was concerned with the assessment of the independent and interactive effects of social class, age, and type of reinforcement on the operant conditioning of verbal behavior of children. Consistent with predictions, the lower-class second-grade subjects were significantly more responsive to material than to social and symbolic reinforcement, but the difference between the effects of social and symbolic reinforcement was not significant. Contrary to expectations, the middle-class second-grade subjects evidenced no significant differences in responsiveness to any of the three types of reinforcement. At the sixth-grade level, lower-class subjects showed no significant differences in their responsiveness to the three types of reinforcement. The middle-class sixth-grade subjects showed no significant difference between the effectiveness of praise and symbolic reinforcement; yet, as expected, praise and symbolic reinforcement each produced a significantly greater increase in response rate than did material reinforcement.

A unique contribution of this study was the opportunity to study the interactive effects of age level, social class, and type of reinforcement. When the scores from second- and sixth-grade subjects were combined, a significant interaction between social class and type of reinforcement was obtained. That is, middle-class subjects were generally more responsive to symbolic and social reinforcement than to material reinforcement, while the lower-class subjects were more responsive to material than to social and symbolic reinforcement.

When the data from both social classes were combined, there was a significant interaction between grade level and type of reinforcement. The older subjects were generally more responsive to social and symbolic reinforcement than to material reinforcement, while the younger subjects had a tendency to be more responsive to material than to social and symbolic reinforcement.

The finding that lower-class children in the second grade were more responsive than second-grade middle-class subjects to material reinforcement and the greater responsiveness of middle-class sixth-grade subjects than lower-class sixth-grade subjects to praise and symbolic reinforcement is consistent with the results of the previously discussed studies by Terrell et al. (1957, 1959), Zigler et al. (1962), Staats et al. (1965), and Cameron et al. (1965). Contrary to previous findings by Zigler and Kanzer (1962), this study did not show a significant difference between the effectiveness of praise and symbolic reinforcement for either lower- or middle-class subjects at either Grade 2 or 6.

The finding that middle-class sixth graders were more responsive to symbolic reinforcement than second graders is congruent with the findings of Lewis et al. (1963). This trend was not evidenced among lower-class children, as the increase in value placed on social or symbolic reinforcement with an age increase was not significant. The latter finding is inconsistent with the significant interaction between grade level and type of reinforcement for lower-class children that was obtained in an earlier study (Cradler, 1968). However, it should be noted that the subjects employed in the earlier study were of a lower socioeconomic status (unskilled or unemployed) than were the lower-class subjects (skilled manual employees and semi-skilled employees) used in the present study.

An additional finding of the present study was that subjects aware of the reinforcement contingency conditioned more rapidly and/or to a greater degree than unaware subjects. Awareness of reinforcement contingencies was more evident for sixth- than second-grade subjects of both social classes and more evident for middle- than lower-class subjects at both age levels.

An interpretation of these findings is that the reinforcer served not only to condition subjects to the word "they," but to motivate them to concentrate on trying to determine the correct contingency between the reinforcement and their own responses.

The lack of statistical support of some of the hypotheses may have been due to the large standard deviations for most of the 12 cells or groups. Therefore, in future research, error variance might be lessened by (*a*) providing a clearer distinction between social classes, (*b*) employing a task which is less affected by extraneous variables than the Taffel (1957) verbal conditioning task, (*c*) controlling for the sex of the experimenter, and (*d*) making it less likely that the subjects will become aware of the reinforcement contingencies.

REFERENCES

BAER, P. E., & GOLDFARB, G. E. A developmental study of verbal conditioning. *Psychological Reports,* 1962 **10,** 175–181.

CAMERON, A., & STORM, T. Achievement motivation in Canadian Indian, middle and working-class children. *Psychological Reports,* 1965, **16,** 459–463.

CRADLER, J. D. The effects of age and type of reinforcement on the operant conditioning of verbal behavior of lower-class children. Unpublished manuscript, San Jose State College, 1968.

DOUVAN, E. Social status and success strivings. *Journal of Abnormal and Social Psychology,* 1956, **52,** 219–265.

HEROLD, P. L. Effects of reinforcement and examiner variables upon persistence among disadvantaged Mexican-American kindergarteners. Unpublished manuscript, San Jose State College, 1967.

HOLLINGSHEAD, A. B., & REDLICH, F. F. *Social-class and mental illness.* New York: Wiley, 1958.

JEFFREY, W. E., & SKAGER, R. W. Effects of incentive conditions on stimulus generalization in children. *Child Development,* 1962, **33,** 865–870.

KATZ, I. Some motivational determinants of racial differences in intellectual achievement. *International Journal of Psychology,* 1967, **2,** 1–12.

LEWIS, M., WALL, M., & ARONFREED, J. Developmental change in the relative values of social and nonsocial reinforcement. *Journal of Experimental Psychology,* 1963, **66,** 133–137.

RISLEY, T. Learning and lollipops. *Psychology Today,* 1968, **8,** 28.

ROSENHAN, D. L. Effects of social class and race on responsiveness to approval and disapproval. *Journal of Personality and Social Psychology.* 1966, **4,** 253–259.

ROSENHAN, D. L., & GREENWALD, J. A. The effects of age, sex, and socioeconomic class on responsiveness to two classes of verbal reinforcement. *Journal of Personality,* 1965, **33,** 108–121.

SPIELBERGER, C. D., LEVIN, S. M., & SHEPARD, M. The effects of awareness and attitude toward the reinforcement on the operant conditioning of verbal behavior. *Journal of Personality,* 1962, **30,** 106–121.

STAATS, A. W., & BUTTERFIELD, W. H. Treatment of nonreading in a culturally deprived juvenile delinquent: An application of reinforcement principles. *Child Development,* 1965, **36,** 925–942.

SULLIVAN, P. W. The effects of verbal reward and verbal punishment on concept elicitation in children. *American Psychologist,* 1960, **15,** 401.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *Journal of Abnormal and Social Psychology,* 1955, **51,** 496–501.

TERRELL, G., DURKIN, K., & WIESLEY, M. Social class and the nature of the incentive in discrimination learning. *Journal of Abnormal and Social Psychology,* 1959, **59,** 270–272.

WALTERS, R. H., & FOOTE, A. A study of reinforcer effectiveness with children. *Merrill-Palmer Quarterly,* 1962, **8,** 149–157.

WITRYOL, S. L., & ORMSBY, E. L. Age trends in children's incentives scaled by paired comparisons. In S. L. Witryol, D. J. Tyrrell, & L. M. Lowden (Eds.), Development of incentive values in childhood. *Genetic Psychology Monographs,* 1965, **72,** 201–246.

ZIGLER, E., & DELABRY, J. Concept switching in middle-class, lower-class and retarded children. *Journal of Abnormal and Social Psychology,* 1962, **60,** 267–273.

ZIGLER, E., & KANZER, P. The effectiveness of two classes of verbal reinforcers on the performance of middle- and lower-class children. *Journal of Personality,* 1962, **30,** 157–163.

(Received March 13, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 4, 285–289

EFFECTIVENESS OF STUDENT-TO-STUDENT COUNSELING ON THE ACADEMIC ADJUSTMENT OF POTENTIAL COLLEGE DROPOUTS†

WILLIAM F. BROWN, NATHAN O. WEHE, VERNON G. ZUNKER
Southwest Texas State University

AND WARREN L. HASLAM
Laredo Junior College

One hundred and twenty-four beginning college freshmen identified as potential dropouts were provided academic adjustment counseling by carefully selected and trained upperclassmen at Southwest Texas State University during the fall of 1967. One hundred and eleven of the 124 students receiving the counseling were individually matched with a control group of 111 potential dropouts that were denied the counseling. Matching of the two groups was done on the basis of age, sex, American College Test composite score, high school quarter rank, and high school size. All students resided in University residence halls. Following the counseling, students in the experimental and control groups were compared on four indexes of counseling results: (*a*) pre–post counseling scores on the Survey of Study Habits and Attitudes, (*b*) pre–post counseling scores on the Effective Study Test, (*c*) post-counseling scores on the Study Skills Surveys, and (*d*) first-semester grade averages. The experimental group was found to be significantly higher on all four criterion measures.

Constantly increasing college enrollments have forced many 4-year institutions to initiate selective admissions programs and thereby reduce the percentage of their students having questionable potential for academic success. Nevertheless, the national trend toward ready availability of higher education for all suggests the freshman year attrition rate is likely to remain at about 40%. Most institutions, therefore, recognize the need for expanding their counseling activities in an effort to improve the academic adjustment of their beginning freshmen. However, lack of financial resources and trained personnel have generally restricted the expansion of professional counseling services on the college campus. In response, many colleges are turning to the use of carefully selected and trained upperclassmen as a means of augmenting their counseling programs. In fact a recent survey indicates a significant increase in the utilization of upperclassmen as first-echelon counselors at senior colleges and universities (Zunker & Brown, 1966).

Research during the past few years has demonstrated that student-to-student counseling on academic-adjustment problems is both a practical and economical means for increasing the chances of scholastic success for the average college student (Brown, 1965a). Further research indicates that carefully selected and trained student counselors are as effective and as accepted as are professional counselors in providing academic-adjustment counseling for the typical college freshman (Brown & Zunker, 1966).

A survey of the research literature, however, indicates a general lack of information about the effectiveness of such student-to-student counseling for freshmen identified as potential college dropouts. This study was, therefore, designed to evaluate systematically the effectiveness of student-to-student counseling upon the scholastic adjustment and subsequent academic survival of college freshmen identified as potential dropouts.

---

[1] This study was supported by a grant from the ESSO Education Foundation, New York, New York. It was based, in part, on an unpublished master's thesis by one of the authors, Nathan O. Wehe, submitted to Southwest Texas State University in 1967.

[2] Requests for reprints should be sent to William F. Brown, P.O. Box 603, San Marcos, Texas 78666.

METHOD

The research project was specifically designed to evaluate the effectiveness of a specially structured student-to-student counseling program providing academic-adjustment guidance to college freshmen identified as potential dropouts. More specifically, the counseling program was evaluated as to its impact upon the scholastic motivation, study behavior, and subsequent academic achievement of potential college dropouts.

Experimental students were selected and received academic-adjustment counseling whereas an individually matched control group was denied such counseling. Upon completion of counseling, the two matched groups, each containing 111 students, were compared on four indexes of counseling results—study orientation, study knowledge, study problems, and semester grade point averages. In addition, the effectiveness and acceptability of the counseling program was evaluated by the experimental group.

Program Description

The objective of student-to-student counseling was to increase the probability of scholastic survival for beginning freshmen identified as potential college dropouts. The student academic counselors were systematically trained to accomplish the following activities during counseling: (*a*) to employ peer-group discussion procedures, (*b*) to communicate academic survival information, (*c*) to disseminate standardized test results, (*d*) to analyze potential academic and social adjustment problems, and (*e*) to initiate individualized corrective measures for identified problems.

As reported in Table 1, the program consisted of four sequential counseling activities. The first activity, Survival Orientation, was held during freshman orientation week and consisted of a 2-hour large group ($n = 30$) meeting in which the student counselors reported on common adjustment problems facing beginning college freshmen. Attendance at this first session was required of all beginning freshmen, whereas participation in the remaining sessions was on a voluntary basis. The second counseling activity, Test Interpretation, was held during the first 4 weeks of the semester and consisted of a 2-hour small group ($n = 4$) session in which the students discussed various factors determining scholastic success. The third activity, Study Skills Instruction, was conducted during the next 4 weeks of the semester and provided individualized assistance in one or more of six problem areas—budgeting time, taking class notes, reading textbook assignments, writing themes and reports, remembering and reviewing materials, and taking essay and objective tests. The fourth and final counseling activity, Study Habits Evaluation, was held during the tenth week of the semester. This final individualized session was designed to provide additional adjustment counseling for those students who received midterm failing notices. Table 1 summarizes the counseling activities provided.

Subjects

Freshmen who entered Southwest Texas State University with a composite standard score of 18 or below on the American College Test (ACT) were considered to be potential dropouts. Institutional research had demonstrated that freshmen who score 18 and below have only one chance in three of continuing their second year in good standing. Experimental and control subjects were, therefore, limited to students who attained an ACT composite standard score of 18 or below and were first-semester freshmen.

Procedure

All entering freshmen were required to attend a 2-hour Survival Orientation session as part of the University's general orientation program. Students met in groups of 30 and were counseled about general college adjustment problems by two student counselors. At each group session all freshmen completed an information form giving their name, age, sex, high school size, high school quarter rank, enrollment classification, and college address. Student counselors utilized the Testing and Guidance Center files to verify the recorded information and to add to ACT composite scores to each information form.

At the conclusion of Survival Orientation, all students with an ACT composite of 18 or below were encouraged to register for additional academic-adjustment counseling. The experimental group was selected from those potential dropouts who voluntarily participated in the postorientation activities of the counseling program. A like number of freshmen were selected from a control group of identified potential dropouts who had also volunteered for the additional counseling but who were subsequently unable to participate in the project. The experimental and control subjects were individually matched on age, sex, ACT composite scores, high school quarter rank, and high school size. All subjects resided in university residence halls. From the 124 potential dropouts receiving the additional counseling, 111 students (experimental group) were individually matched with 111 students (control group) that were denied the additional counseling. Comparison of precounseling and postcounseling scores on the Survey of Study Habits and Attitudes (Brown & Holtzman, 1967) and the Effective Study Test (Brown, 1964), and postcounseling scores on the Study Skills Surveys (Brown, 1965b) were used to evaluate the effectiveness of the counseling program. First-semester grade point averages were compared to evaluate the counseling program's productivity. At the conclusion of the student-to-student counseling activities, critical aspects of the counseling program were evaluated by the experimental group through the use of a specially constructed Counseling Evaluation Questionnaire.

TABLE 1

GUIDANCE ACTIVITY SEQUENCE

Session	Time period	Group size	Activity description	Duration
1	Freshman orientation week	30	*Survival Orientation:* Briefing session on campus community organization, college survival facts, academic-adjustment problems, scholastic achievement standards, graduation requirements, time management principles, environmental variables influencing study efficiency, and student assistance resources.	2 hours
2	First–fourth weeks	4	*Test Interpretation:* Counseling session on factors determining scholastic success, results of standardized ability and achievement tests, current study habits and attitudes, common personal–social adjustment problems, factors influencing peer-group affiliation, and college rewards and frustrations.	2 hours
3–8	Fifth–eighth weeks	1	*Study Skills Instruction:* Counseling session on how to read and remember textbook material, take lecture and reading notes, write themes and reports, prepare for essay and objective examinations, and organize study time and place.	1 hour each topic
9	Tenth week	1	*Study Habits Evaluation:* Counseling session on probable causes for unsatisfactory midsemester grade reports and appropriate corrective action for residual study organization, techniques, and motivation problems.	1 hour

Materials

The instructional package for the specially structured counseling program included a student counselor's handbook, a student's textbook, assorted illustrated handouts, discussion questionnaires, overhead transparencies, and other supporting guidance aids developed by Brown and Holtzman (1964). These materials were used to discuss such topics as taking lecture notes, taking reading notes, organizing study environment, reading textbook assignments, organizing study time, writing in-class themes, preparing outside reports, giving oral reports, remembering specific details, understanding broad concepts, preparing for examinations, taking examinations, etc. In addition, materials on seeking assistance, facing reality, accepting responsibility, and motivating study were also utilized during the counseling activities.

A specially constructed Student Counselor's Handbook (Brown, 1967) served as a guide for the student counselors in using the counseling materials. The handbook provides a detailed description of the systematic counseling program and procedures employed in this study.

RESULTS AND DISCUSSION

The Survey of Study Habits and Attitudes was administered to the experimental and control groups both before and after the experimental group received counseling. Comparisons of test–retest differences in the scores earned by the two groups provided one indication of the effectiveness of the additional counseling received by the experimental group. The Survey of Study Habits and Attitudes is an easily administered measure of study methods, motivation for studying, and certain attitudes toward scholastic activities which influence classroom success. Table 2 reports the means and standard deviations for the test–retest difference scores earned by the two groups in the research sample. Fisher's t test for determining the significance between correlated means was computed and the results are also reported in Table 2. These data suggest that members of the experimental group experienced a significant positive change in their study orientation whereas the study orientation of the control group members remained relatively unchanged.

Test–retest difference scores from the Effective Study Test were used to compare changes in knowledge about efficient study

TABLE 2

COMPARISON OF EXPERIMENTAL (COUNSELED) AND CONTROL (UNCOUNSELED) FRESHMEN ON FOUR CRITERION MEASURES

Measure	Experimental group		Control group		M_{diff}	t
	M	SD	M	SD		
SSHA difference score[a]	14.40	30.03	.13	27.11	14.27	3.87**
EST difference score[a]	2.68	8.03	−.60	9.73	3.28	2.69**
SSS score	25.10	9.50	27.70	8.71	−2.60	2.15*
First semester GPA	1.77	.72	1.52	.70	.25	3.13**

Note.—$n = 111$ for each group. Abbreviated: SSHA = Survey of Study Habits and Attitudes, EST = Effective Study Test, SSS = Study Skills Surveys, GPA = grade point average.

[a] Test–retest difference score computed by subtracting initial score from final score.

* $p < .05$.

** $p < .01$.

practices. The Effective Study Test is designed to measure a student's knowledge about efficient study methods and the factors influencing their development. Table 2 also reports the means and standard deviations for the test–retest difference scores earned by the two groups, together with the results for Fisher's t test for determining the significance between correlated means. These data indicate that members of the experimental group significantly increased their knowledge about effective methods of study whereas members of the control group remained relatively unchanged in their study skills knowledge.

Comparisons of postcounseling scores on the Study Skills Survey for the experimental and control groups provided additional data for assessing changes in study organization, techniques, and motivation. The Study Skills Survey is designed as a self-report procedure in which the individual indicates his existing study problems. Table 2 also gives the means and standard deviations for the experimental and control samples, together with the results for Fisher's t test between correlated means. From Table 2 it may be seen that there was a significant difference in self-reported study problems in favor of the experimental group. These results indicate that students receiving the additional counseling not only improved their study procedures and increased their how-to-study knowledge, but also reported significantly fewer study skills problems than did students who were denied the additional counseling.

The counseling program's productivity was evaluated through a comparison of first-semester grade point averages. Grade point averages were calculated on the basis of 3, 2, 1, 0, and 0 points for the letter grades of A, B, C, D, and F, respectively, Fisher's t test for correlated means was again used to determine level of significance. As reported in Table 2, these data reveal that there is a significant difference in earned grades in favor of the student group that received the additional counseling.

The Counseling Evaluation Questionnaire was anonymously administered to the experimental students after completion of the counseling program. The questionnaire was designed to determine the counselee's acceptance or rejection of the student-to-student approach, the counseling materials and procedures, and the counseling objectives and results. Five possible responses, ranging from strongly agree to strongly disagree, were provided for the counselee to indicate his degree of agreement or disagreement with each of the questionnaire's 60 statements. Counselee responses were assigned a weighted score of +2 for very positive through −2 for very negative. Tabulation of responses to all 60 statements indicated that the experimental students' reactions were decisively positive to all evaluated aspects of the program. Furthermore, the strongest aspect of the counseling program, as perceived by the counselees themselves, was the student-to-student approach, while the weakest perceived aspect of the program was the lack of sufficient time for comprehensive discussion of identified study skills problems. These two factors appear to complement each other, for the counselees would apparently want more time to discuss their problems with student counselors only if such counseling were perceived as helpful by the counselees.

Several specific conclusions appear warranted by the research results. First, potential dropouts receiving the additional student-to-student counseling showed significant positive changes in their measured study orientation and study knowledge, as indicated on the Survey of Study Habits and Attitudes and Effective Study Test, whereas potential dropouts not receiving such counseling did not show such improvement. Second, the counseled group reported fewer residual study problems, as indicated on the Study Skills Survey, than did the group which received only the orientation counseling. In fact, students receiving additional counseling beyond the orientation counseling showed significant improvement on all variables measuring effective study procedures and scholastic motivation. Third, students receiving the additional counseling made significantly better grade averages than did matched students not receiving such help, suggesting that the counselees were putting the information received during counseling to effective use. Finally, the counselee's evaluation of the counseling program, as indicated on the Counseling Evaluation Questionnaire, was decisively positive on all evaluated aspects of the counseling program.

It can be concluded that student-to-student counseling designed to improve study skills, scholastic motivation, and academic achievement is a practical and productive means for aiding the student whose measured potential for successfully completing college is questionable.

REFERENCES

BROWN, W. F. *Effective Study Test: College level,* San Marcos, Tex.: Effective Study Materials, 1964.

BROWN, W. F. Student-to-student counseling for academic adjustment. *Personnel and Guidance Journal,* 1965, **43,** 811–817. (a)

BROWN, W. F. *Study Skills Surveys.* San Marcos, Tex.: Effective Study Materials, 1965. (b)

BROWN, W. F. *Student counselor's handbook.* San Marcos, Tex.: Effective Study Materials, 1967.

BROWN, W. F., & HOLTZMAN W. H. *Effective Study Course.* San Marcos, Tex.: Effective Study Materials, 1964.

BROWN, W. F., & HOLTZMAN, W. H. *Survey of Study Habits and Attitudes: College level.* New York: Psychological Corporation, 1967.

BROWN, W. F., & ZUNKER, V. G. Student counselor utilization at four-year institutions of higher learning. *Journal of College Student Personnel,* 1966, **64,** 41–46.

ZUNKER, V. G., & BROWN, W. F. Comparative effectiveness of student and professional counselors. *Personnel and Guidance Journal,* 1966, **64,** 738–743.

(Received September 30, 1969)

Journal of Educational Psychology
1971, Vol. 62, No. 4, 290–292

EFFECTS OF STRUCTURAL COMPLEXITY OF BIGRAMS ON PAIRED-ASSOCIATE LEARNING AND RETENTION OF CHILDREN[1]

ROBERT A. HICKS[2] AND JAMES PACKWOOD
San Jose State College *University of Oregon*

The verbal learning and retention of 36 first- and second-grade children was measured as a function of the structural complexity of the letters used. Specifically, a list of six pairs of bigrams composed of two high-, two moderate-, and two low-complexity pairs was learned, then relearned 7 days later. Complexity had a highly significant effect upon rate of learning. However, these data only partially supported the hypothesis of an inverse relationship between level of complexity and trials to learn, in that the effect was produced primarily by the relatively greater ease with which the high-complexity bigrams were mastered. The effect of complexity on relearning was even more pronounced and, further, these data showed, in support of the hypothesis, a significant inverse relationship between complexity and trials to relearn.

Berlyne (1960) has postulated that complexity is a stimulus attribute that serves to enhance level of arousal and thereby effect attention, that is, given two stimuli which vary in degree of complexity, an individual is likely to attend to the relatively more complex stimulus. There is support for this hypothesis. The direct relationship postulated between stimulus complexity and behavioral measures of attention has been well documented. Several studies, for example, Berlyne & Slater (1957), Berlyne (1958a, 1958b), Frantz (1958), and Cantor, Cantor, & Ditrichs (1963) have shown that both human and infrahuman subjects, when given a choice, spend more time looking at the more complex member of a set of stimuli.

Recently, Berlyne (1965) has shown, consistent with an aspect of the aforementioned hypothesis, that stimulus complexity has a direct effect on electrophysiological indexes of level of arousal. As Easterbrook (1959), among others, has pointed out, increases in level of arousal serve to focus attention. Hence it seems justified to conclude that degree of attention to a stimulus is, in part, a function of the complexity of that stimulus.

Bugelski (1964) has suggested that attention is a critical variable in learning and that any variable that improves attention should enhance learning and retention which involves that variable. Butler (1964) measured the effects of stimulus complexity on learning lists of symbolic digits. Her data suggest a direct relationship between stimulus complexity and learning proficiency. There seems, then, sufficient reason to believe that stimulus complexity should be a significant factor in verbal learning and retention. The purpose of this study was to investigate this relationship. Specifically, it was postulated that trials to criterion in paired-associate verbal learning and relearning of bigrams would be inversely related to the level of complexity of the verbal units used, with complexity defined by the structure of the letters in each bigram.

METHOD

Subjects

The subjects were 36 first- and second-grade students who were selected so that each knew the alphabet.

Apparatus

The subjects were tested using a learning apparatus which was similar in many respects to one used by Brackbill and Kappy (1962). Essentially the apparatus consisted of a gray plywood sheet on which was mounted two apertures, one white the other black, which were covered with sheets of one-way glass. The stimulus and response terms were presented in these. Hunter timers were wired so that they controlled presentation time by turning off and on a 40-watt high-intensity bulb that was housed above each aperture thereby illuminating the bigrams so that these could be seen by the subject.

---

[1] These results were presented at the meetings of the Western Psychological Association, Vancouver, British Columbia, June 1969.

[2] Requests for reprints should be sent to Robert Hicks, Department of Psychology, San Jose State College, 125 South Seventh Street, San Jose, California 95114.

Stimuli

The bigrams were printed in 2-inch black letters on white plywood strips. The structural complexity of each letter was determined by comparing that letter to a generation matrix used by Alluisi and Martin (1958) and Hughes (1961). The letters which included curved lines were made angular to better fit the matrix. The complexity of each bigram was determined by summing the complexity values (i.e., number of elements) of its two letters. In this manner, a set of six pairs of bigrams was generated which was composed of two low-complexity pairs, LU-TC, UT-NL; two moderate-complexity pairs, FY-DH, AF-YD; and two high-complexity pairs, WS-GM, RB-EG. In constructing this set of bigrams the following considerations were taken into account: (*a*) Frequency of usage of the bigrams was equated across subsets using tables provided by Underwood and Schulz (1960); (*b*) a given letter was used only once in the first or second position of a bigram; (*c*) a given letter appeared only once in a given stimulus-response pair. Finally it should be noted that the use of a list composed of six pairs of bigrams was determined on the basis of pilot research as the longest feasible list for first and second graders.

Procedure

Each subject was brought to the test room and for each subject at least 5 minutes was taken to establish rapport. The instructions were then read to the subject. Understanding of these instructions was tested using a practice paired-associate bigram pair. After the experimenter was certain that the subject understood the task, the stimuli were presented. On the basis of pilot research, it was determined that the stimulus bigram be presented alone for 4 seconds, then paired with the response bigram for 2 seconds with a 10-second intertrial interval. The order of presentation of each pair was counterbalanced so that each pair occurred equally as often under each stage of practice. Further, for half the subjects the stimulus-response terms were inverted. The dropout method was used, that is, when a subject reached criterion of three errorless trials in a row for a given stimulus-response pair, that pair was dropped from the list. After 7 days each subject returned and relearned the list using the same order utilized initially during learning. Trials to criterion was taken as the measure of performance for each bigram. Scores for high, medium, and low complexity were the summated trials to criterion for the two bigrams representing each of these dimensions.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF SUMMATED TRIALS TO CRITERION FOR THE HIGH-, MEDIUM-, AND LOW-COMPLEXITY BIGRAMS

Bigram complexity	Acquistion		Relearning	
	M	*SD*	*M*	*SD*
High	15.89	4.89	9.14	2.45
Medium	27.56	8.06	14.28	4.92
Low	25.44	7.69	16.72	4.52

RESULTS

The means and standard deviations for learning and relearning are given in Table 1. These data were tested using one-way analyses of variance for repeated measures designs (Winer, 1964). The effects of complexity on both learning and relearning were found to be significant; for learning, $F = 57.51$, $df = 2/70$, $p < .001$ and for relearning, $F = 91.88$, df 2/70, $p < .001$. The differences between the individual treatment means for both learning and relearning were tested using the direct difference method of computing t and these results are given in Table 2.

DISCUSSION

In general, the results tend to support the predicted direct relationship between degree of complexity and rate of acquisition and level of retention. During acquisition the high-complexity stimuli, as shown in Table 2, were learned in significantly fewer trials than either the low- or medium-complexity stimuli. However, contrary to prediction, the low- and medium-complexity stimuli

TABLE 2

t TESTS OF THE DIFFERENCES BETWEEN INDIVIDUAL TREATMENT MEANS FOR ACQUISITION AND RELEARNING

Complexity level	Acquisition		Relearning	
	High	Medium	High	Medium
Medium	10.89		8.29	
Low	8.39	−1.73[a]	13.30	3.84

Note.—$t_{.05} = 2.03$, $df = 35$.

[a] Minus (−) value indicates superiority of the row variable.

did not differ significantly in trials to learn. In fact, as shown in Table 1, the medium bigrams took more trials to learn than did the low-complexity bigrams.

As shown in Table 1, and consistent with the hypothesis, there is a direct relationship between stimulus complexity and degree of retention. Further, as shown in Table 2 each increment in complexity is associated with a significant increase in retention.

These explanations can be advanced to explain the lack of conformity to the hypothesis for the low- and medium-complexity bigrams during acquisition. First, the matrix used made it possible to generate bigrams on a complexity continuum that varied only from 3 to 10 elements. It is possible that the low- and medium-complexity verbal units were not sufficiently different to produce a significant change in behavior. However, this possibility can be questioned by the significant effects found during relearning. It seems more probable that, since the differences between the low and medium lists were not significant, the direction of these differences could be considered a chance fluctuation. Taken together with the retention data, the acquisition data suggest that complexity exerts a greater direct effect on retention. From these data it seems obvious to conclude that future studies of verbal learning, which use children as subjects must consider list complexity as a variable. Logically the effects demonstrated here are probably at a maximum for first and second graders. The complexity of letters, as would the perceived complexity of any stimulus, diminishes with experience. Hence, the structural complexity of letters might not have an effect on the verbal learning of adults. However, these age parameters should be determined empirically.

REFERENCES

Alluisi, E. A., & Martin, H. B. An information analysis of verbal and motor responses to symbolic and conventional arabic numerals. *Journal of Applied Psychology,* 1958, **42,** 79–84.

Berlyne, D. E. The influence of complexity and novelty in visual figures on orienting responses. *Journal of Experimental Psychology,* 1958, **55,** 289–296. (a)

Berlyne, D. E. The influence of the albedo and complexity of stimuli on visual fixation in the human infant. *British Journal of Psychology,* 1958, **49,** 315–318. (b)

Berlyne, D. E. *Conflict, arousal, and curiosity.* New York: McGraw-Hill, 1960.

Berlyne, D. E. Effect of stimulus complexity and incongruity on duration of EEG desynchronization. *Electroencephalography and Clinical Neurophysiology,* 1965, **18,** 156–161.

Berlyne, D. E., & Slater, J. Perceptual curiosity, exploratory behavior and maze learning. *Journal of Comparative and Physiological Psychology,* 1957, **50,** 228–232.

Brackbill, Y., & Kappy, M. S. Delay of reinforcement and retention. *Journal of Comparative and Physiological Psychology,* 1962, **55,** 14–18.

Bugelski, B. R. *The psychology of learning applied to teaching.* New York: Bobbs Merrill, 1964.

Butler, F. The role of complexity in the recognition of symbolic digits. Unpublished master's thesis, University of Denver, 1964.

Cantor, G. N., Cantor, J., & Ditrichs, R. Observing behavior in pre-school children as a function of stimulus complexity. *Child Development,* 1963, **34,** 683–689.

Easterbrook, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychological Review,* 1959, **66,** 183–201.

Frantz, R. L. Pattern vision in young infants. *Psychological Record,* 1958, **8,** 43–48.

Hughes, C. L. The variability of stroke-width within digits. *Journal of Applied Psychology,* 1961, **45,** 364–368.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

Winer, R. B. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received November 15, 1969)

Journal of Educational Psychology
1971, Vol. 62, No. 4, 293-299

EFFECTS OF INTELLIGENCE, INFORMATION PROCESSING, AND MEDIATION CONDITIONS ON CONCEPTUAL LEARNING[1]

JIM MILLHAM, LEONARD I. JACOBSON,[2] AND STEPHEN E. BERGER
University of Miami, Coral Gables

The effects of associative and inferential information-processing ability and mediation conditions on conceptual learning were investigated. The two conceptual tasks employed promoted the differential use of symbolic processes. The massed practice task facilitated symbolic information processing and minimized interference effects; the distributed practice task disrupted information processing and promoted interference effects. Both tasks included mediation and nonmediation conditions. It was found that the 231 undergraduate volunteers learned concepts most effectively under massed practice and mediation conditions and least effectively under distributed practice and nonmediation conditions. Superior performance under massed practice resulted from subjects scoring high in associative information-processing ability. Subjects scoring high in inferential information-processing ability learned concepts more rapidly than lower scorers solely in the mediation condition of the massed practice task.

The present study investigated the effects of two measures of intelligence, Vocabulary and Abstraction, conceptualized as indicators of associative and inferential information processing, respectively, and mediation conditions on the learning of two conceptual tasks that promoted the differential use of symbolic processes.

In previous studies of the relationship of intelligence to conceptual learning, it has been demonstrated that performance is contingent upon the complex interaction of subject variables and the organization and structure of learning tasks (Jacobson, Dickinson, Fleishman, & Haraguchi, 1969; Jacobson, Elenewski, & Lordahl, 1968; Jacobson, Elenewski, Lordahl, & Liroff, 1968; Jacobson, Millham, & Berger, 1969; Jensen, 1966; Osler & Fivel, 1961; Osler & Trautman, 1961; Osler & Weiss, 1962).

These investigations have identified specifically and demonstrated experimentally the critical importance of each of the following factors: (*a*) the individual difference variable investigated (e.g., intelligence or creativity); (*b*) the particular measure of a variable used (e.g., Wechsler Adult Intelligence Scale or Shipley-Hartford intelligence test); (*c*) the learning task employed (e.g., rote or concept learning); (*d*) the particular paradigm used (e.g., mediation or nonmediation paradigm); (*e*) the performance indicator employed (e.g., trials or errors); (*f*) the information provided by the task (e.g., amount of relevant or irrelevant information provided); and (*g*) the stage of the learning process investigated (e.g., period to the first correct response or period from the first correct response to concept attainment).

The results obtained have been found to vary as a function of the interaction of experimental modifications in each of these factors. For example, Osler and Trautman (1961) found that subjects of high intelligence demonstrated more effective concept learning than other subjects on tasks in which information was provided that was useful in symbolic information processing. However, the advantage displayed by high-intelligence subjects was not found in a similar concept-learning situation in which the amount of irrelevant information was increased.

---

[1] The authors express their thanks to Abby Deutsch, Frederick L. Newman, and Jorge A. Siblesz for their assistance. This study was supported by Research Grant NGL 10-007-010 from the National Aeronautics and Space Administration.

[2] Requests for reprints should be sent to Leonard I. Jacobson, Department of Psychology, University of Miami, P.O. Box 8185, Coral Gables, Florida 33124.

In another study (Jacobson, Dickinson, Fleishman, & Haraguchi, 1969), intelligence was measured by the School and College Abilities Test (SCAT), a test of general intelligence. It was found that subjects of high intelligence learned both mediated and nonmediated concepts more rapidly than subjects of lower intelligence. No significant interaction of intelligence and mediation conditions was in evidence.

However, in a second study (Jacobson, Millham, & Berger, 1969), intelligence was measured by scores on the Vocabulary test of the Wechsler Adult Intelligence Scale (WAIS). The procedure and materials employed in this investigation were generally similar to those in the study of Jacobson, Dickinson, Fleishman, & Haraguchi (1969). In contrast to the earlier study, it was found that intelligence interacted with mediation conditions. Subjects of high intelligence learned mediated concepts more rapidly than subjects of lower intelligence, but this effect was not found in the learning of nonmediated concepts. Thus, the two intelligence measures interacted differentially with the experimental conditions.

The concept-learning procedure employed in these studies was the method of successive presentation often referred to as the response method. In this procedure, subjects are presented with a series of stimuli and are required to name each stimulus as it appears as an instance of a particular concept.

Conceptual learning based on this procedure may be divided into two stages (Jacobson, Elenewski, Lordahl, & Liroff, 1968). The first stage, the period until the first correct response, has been termed the response-generation stage of conceptualization. During this period the problem is a generative one and subjects must process information effectively in order to generate the first correct response.

The period from the first correct response until concept attainment has been termed the recognition stage of conceptualization. During this period, the subject must generalize a previously attained correct response to related instances in a set of a limited size. Previous studies have indicated that the effects of both mediation and intelligence occur during the response-generation stage but not during the recognition stage of conceptualization (Jacobson, Dickinson, Fleishman, & Haraguchi, 1969; Jacobson, Elenewski, Lordahl, & Liroff, 1968; Jacobson, Millham, & Berger, 1969).

These studies indicate that intelligence and mediation conditions have been experimentally related to symbolic information processing in concept learning. However, the conflicting results obtained as a result of the interaction of intelligence measures and experimental tasks and conditions indicate that additional investigations are necessary in order to clarify the intelligence construct.

The present study investigated the relationship of two variables that measure differential aspects of intelligence to the learning of different concept materials and procedures. The two tasks employed in the present study were both variants of the method of simultaneous presentation which is the most widely used alternative to the response method of concept learning.

In the method of simultaneous presentation all instances of a concept are presented together as a set and the subject attempts to identify simultaneously all of the elements as instances of the correct concept. Concept attainment in this procedure is often assumed to be greatly contingent on the use of symbolic processes, in contrast to the response method which is frequently viewed as a variant of rote learning (Bourne, 1966; Deese & Hulse, 1967).

Two variants of the method of simultaneous presentation were employed. In the massed practice variant, the elements of each concept were presented consecutively for 20 trials; in the distributed practice variant each set of elements was presented for a total of 20 trials but the sets were presented in mixed rather than in consecutive order.

On the massed practice task, the subject was required to learn one concept at a time and relevant instances of the concept were presented simultaneously in consecutive trials. This procedure facilitated the use of symbolic information processing and minimized interference effects. In contrast, the distributed practice task required the subject to attend to three sets of different concept instances and to attempt their solution simultaneously. The procedure was designed to disrupt continuous, symbolic information processing and to promote interference effects. On both tasks, a mediation and a nonmediation condition were employed. Thus, the experimental tasks were arranged so that the relationship of intelligence to the continual use of symbolic information processing and to mediation conditions could be evaluated.

Two measures of intellectual functioning were employed. These were the Vocabulary test of the WAIS and the Abstraction scale of the Shipley-Hartford intelligence test. Research on intelligence indicates that the best single predictor of full scale IQ on the most widely used intelligence tests is the score on the Vocabulary subtest of these scales (Terman & Merrill, 1960; Wechsler, 1949, 1958). WAIS Vocabulary correlates with full scale IQ between .7 and .9 across different age levels.

Vocabulary tests may be viewed as measuring the ability of the subject to respond to a verbal stimulus with a common (conventional) and equivalent (synonymous) associative response. The response is required to be both a common and equivalent associate since it must meet norms found in the dictionary that prescribe those associative responses that are commonly regarded by the language community as stimulus equivalent.

Thus, scores on a vocabulary test may be viewed as measuring associative information-processing ability. Such an interpretation is supported by investigations indicating that vocabulary scores are excellent predictors of the use of associative processes in concept learning employing the response method (Jacobson, Elenewski, Lordahl, & Liroff, 1968; Jacobson, Millham, & Berger, 1969).

In contrast to the interpretation of vocabulary scores as indicators of associative information-processing ability, scores on the Abstraction test of the Shipley-Hartford intelligence scale may be viewed as measures of inferential information-processing ability. As an example, the Abstraction test requires the subject to complete the following line:

"oh ho/ rat tar/ mood. . . ."

Solution of this problem requires the subject to infer the correct rule for problem solution. Unlike the vocabulary test, the generation of a stimulus equivalent to the elements presented is not sufficient. The subject is required to infer, as a result of hypothesis testing, the rule that will solve the problem correctly. Thus, the Abstraction test requires the use of inferential information processing. It is hypothesized that such inferential processing can be conceptually and empirically separated from associative information processing.

METHOD

Subjects and Measures of Intelligence

The subjects were 231 undergraduate volunteers. All of the subjects were administered the Vocabulary test of the WAIS individually, and the Abstraction test of the Shipley-Hartford intelligence scale in a group setting. The reliability of the Vocabulary test at the adult level is approximately .90 (Terman & Merrill, 1960); the reliability of the Abstraction test is .89 (Shipley, 1940). The subjects were divided into high, medium, and low scorers on each test. Subjects scoring in the top third of the distribution for a particular test were considered to be high scorers on that measure. Subjects scoring in the middle and lower thirds of the distribution of scores for a particular test were considered to be medium and low scorers, respectively, on that measure.

Learning Tasks

All of the subjects learned one of two paired-associate lists and a concept list, whose elements were originally constructed by Mednick and Freedman (1960). Both paired-associate lists consisted of 12 word pairs. The first paired-associate list was designed so that the learning of 4 of the 12 word pairs, *sugar-black, hailstone-sheep, button-bread,* and *cabbage-salt,* facilitated the attainment of the concept "white" in later concept learning (Russell & Jenkins, 1954). The remaining word pairs on this list were neutral with regard to "white." The second paired-associate list was designed so as to facilitate the attainment of the concept "soft." On this list it was expected that the learning of four of the word pairs, *pup-hard, grape-loud, peach-smooth,* and *cradle-comfort,* would facilitate the attainment of the concept "soft" on the concept list.

The concept list consisted of three sets of words with four words in each group. The subjects' task was to discover the adjective concept that described each of the four items. The nouns, *sugar, hailstone, button,* and *cabbage,* were instances of the concept "white"; *pup, grape, peach,* and *cradle* were instances of the concept "soft"; *anchor, camel, forest,* and *barrel* were instances of "large."

The attainment of the concept "white," but not of "soft," was facilitated for subjects who learned the first paired-associate list. The attainment of "soft," but not of "white," was facilitated for subjects who learned the second paired-associate list. "Large" was not facilitated on either list and was included as a buffer concept. Thus, possible differences in concept learning as a function of list construction were controlled since both of the experimentally manipulated concepts appeared equally in each condition.

On the massed practice task subjects were presented with the same set of concept instances consecutively for 20 trials. Three orders of this list were employed. On the distributed practice task the three sets of concept instances were each presented for 20 trials but in randomly mixed rather than in consective order.

Mediation and nonmediation concept-learning scores were obtained on both tasks for each subject. The mediation scores were based on the subjects' responses to the items comprising the facilitated concept on the list which he learned. The nonmediation scores were based on the subjects' responses to the items comprising the nonfacilitated concept. Thus, mediation and nonmediation concept scores were obtained for each subject.

Procedure

The subjects learned List 1 or List 2 to a criterion of two successive perfect trials. All subjects then learned the concept materials on either the massed or distributed practice tasks. Three orders of each list for each task were employed. Assignment to paired-associate and concept-learning conditions was randomly determined.

On the massed practice task, the subject learned mediated, nonmediated, and buffer concepts to a criterion of one correct labeling of each concept or were terminated after 20 trials with a concept problem. On the distributed practice task, the subjects were presented with the concept list until a criterion of one correct labeling of each of the three concepts occurred or a total of 60 trials had been completed (20 trials for each concept). Thus, on both tasks subjects had a maximum of 20 trials for each concept group and the tasks differed only in the consecutive or nonconsecutive arrangement of the trials for each concept group. The subjects were then individually administered the WAIS Vocabulary test. The Abstraction scale was administered in a group setting at another time.

The stimuli for the paired-associate lists were presented at 4-second intervals; the stimuli for the concept task were presented at 6-second intervals. All stimuli were presented on a mirascreen projection viewer by means of a slide projector connected to a time interval generator.

RESULTS AND DISCUSSION

Task and Mediation Effects

Table 1 presents the means and standard deviations of the number of trials necessary to attain concepts in the mediation and nonmediation conditions of the massed and distributed practice tasks. The table indicates that concepts were learned more rapidly on the massed practice task than on the distributed practice task. Analysis of variance, presented in Table 2, indicated this difference to be a significant one ($p < .05$).

Table 1 indicates also that concepts were learned more rapidly in the mediation than in the nonmediation condition. Analysis of variance indicated this mediation effect to be a significant one ($p < .001$). Within-task analysis indicated a significant mediation effect on both the massed practice ($F = 4.7$, $df = 1/128$, $p < .03$) and the distributed practice tasks ($F = 4.5$, $df = 1/89$, $p < .04$).

The table indicates that concepts were learned most efficiently in the mediation condition of the massed practice task and least efficiently in the nonmediation condition of the distributed practice task, and this difference was found to be highly significant ($t = 4.2$, $df = 229$, $p < .001$). Although no significant Task × Mediation interaction was obtained, Table 1 indicates

TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE NUMBER OF TRIALS TO LEARN CONCEPTS IN THE MEDIATION AND NONMEDIATION CONDITIONS OF THE MASSED AND DISTRIBUTED PRACTICE TASKS

Task	Mediation		Nonmediation		Across conditions		
	M	*SD*	*M*	*SD*	*M*	*SD*	*n*
Massed	14.5	8.1	16.2	7.7	15.3	7.9	137
Distributed	15.6	7.8	18.3	5.7	16.9	7.0	94

TABLE 2

SUMMARY OF AN ANALYSIS OF VARIANCE OF THE EFFECTS OF VOCABULARY, ABSTRACTION, TASK, AND MEDIATION ON THE NUMBER OF TRIALS TO CONCEPT ATTAINMENT

Source	*df*	*MS*	*F*
Task (T)	1	134.9	4.3*
Vocabulary (V)	2	101.2	3.2*
Abstraction (A)	2	7.5	0.2
T × V	2	57.3	1.8
T × A	2	32.2	1.0
V × A	4	22.8	0.7
T × V × A	4	44.4	1.4
Error (between)	212	31.4	
Mediation (M)	1	488.4	10.4**
T × M	1	30.6	0.7
V × M	2	20.7	0.4
A × M	2	70.9	1.5
T × V × M	2	14.8	0.3
T × A × M	2	167.3	3.6*
V × A × M	4	35.8	0.8
T × V × A × M	4	11.5	0.2
Error (within)	212	46.9	

* $p < .05$.
** $p < .001$.

a mean difference of over two trials to concept attainment between massed practice and distributed practice in the nonmediated learning condition.

Therefore, further analysis was performed and indicated that nonmediated concepts were learned significantly more rapidly on the massed practice than on the distributed practice task ($F = 4.9$, $df = 1/229$, $p < .03$). However, a similar effect was absent in the mediation condition ($F = 1.1$, $df = 1/229$). Thus, the facilitating effects of the massed practice task occurred primarily in the absence of the mediation condition. Both massed practice and mediation were found to facilitate learning, but the joint presence of both conditions did not generate an additive effect.

Vocabulary and Abstraction

Table 2 indicates that a significant effect of Vocabulary was found across tasks ($p < .05$), but no significant interaction of Vocabulary with Mediation, Task, or Abstraction was in evidence. The correlation of Vocabulary and Abstraction was .31 ($N = 231$). Although this correlation is a highly significant one ($p < .001$), as a result of the large sample employed, it should be noted that these variables have less than 10% of their variance in common. Thus, they have a relatively low relationship to each other.

Table 3 presents the means and standard deviations of the number of trials for high, medium, and low Vocabulary scorers to attain mediated and nonmediated concepts on the massed and distributed practice tasks. Table 3 indicates that the general effect of Vocabulary resulted primarily from the differential performance of high, medium, and low Vocabulary scorers on the massed practice task. These specific effects approached significance in both the mediation ($F = 2.1$, $df = 2/128$) and the nonmediation conditions ($F = 3.1$, $df = 2/128$) of massed practice. Thus Vocabulary was significantly related to concept learning and this general effect resulted primarily from performance on the massed practice task.

Further analysis indicated that high Vocabulary scorers learned concepts more rapidly under massed practice than they did under distributed practice. This effect was significant in both the mediation ($t = 2.0$, $df = 74$, $p < .05$), and in the nonmediation conditions ($t = 2.0$, $df = 74$, $p < .05$).

TABLE 3

MEANS AND STANDARD DEVIATIONS OF THE NUMBER OF TRIALS TO LEARN CONCEPTS IN THE MEDIATION AND NONMEDIATION CONDITIONS OF THE MASSED AND DISTRIBUTED PRACTICE TASKS FOR SUBJECTS DIFFERING IN VOCABULARY SCORES

Vocabulary	Massed practice		Distributed practice	
	Mediation	Nonmediation	Mediation	Nonmediation
High				
M	12.8	14.2	16.5	17.5
SD	8.3	8.4	7.7	6.2
n	45	45	31	31
Medium				
M	14.6	16.3	14.7	17.7
SD	7.9	7.6	8.1	6.4
n	43	43	25	25
Low				
M	15.9	18.0	15.4	19.3
SD	7.8	6.6	7.8	4.6
n	49	49	38	38

Journal of Educational Psychology
1971, Vol. 62, No. 4, 300–307

DIFFERENTIAL RESPONSE TO INDUCTIVE AND DEDUCTIVE INSTRUCTIONAL PROCEDURES

MARY LOU KORAN[1]
University of Texas

In an experiment designed to investigate individual differences in learning from inductive and deductive sequences of programmed instruction, 167 subjects were administered aptitude tests representing induction, reasoning, verbal comprehension, and memory. Four sequences of rule and example frames were used to compare inductive and deductive methods, each method having a high- and low-alternation frequency between rules and examples. Criterion performances were speed and accuracy in learning and in stating rules and solving problems on a delayed criterion test. A 2 × 2 analysis of variance disclosed that inductive and deductive methods were about equally effective in time required to teach the information and in promoting transfer, although inductive subjects made significantly more program errors. Regression analyses of Aptitude × Treatment interactions showed that induction, verbal comprehension, and reasoning scores interacted significantly with instructional treatments.

A continuous problem facing educators is that of determining the most effective methods for training large numbers of individuals possessing dissimilar patterns of abilities. A common strategy has been to compare average performance on some criterion in an effort to determine if one method is more effective than others in some general sense. However, learners differ, and the search for generally superior methods should be supplemented by a search for ways in which instruction can be varied so as to fit learner characteristics (Cronbach & Snow, 1969).

The position that learning effectiveness is a function of the interaction of instructional treatments and learner characteristics has both theoretical and empirical support (Cronbach 1957, 1967). A review of recent research (Cronbach & Snow, 1969) suggests that subjects may learn more easily from one method than another, that this best method differs from subject to subject, and that such differences are related to learner characteristics. Accordingly, the incorporation of learner characteristics into learning research and theory appears important for determining more explicitly the individual parameters within which more general principles of learning might be expected to be applicable.

One area in which the role of learner characteristics has not been systematically explored concerns the relative efficacy of inductive and deductive instructional procedures. Despite strong endorsements, a review of research comparing inductive and deductive instructional procedures does not provide consistent empirical support for the use of one or the other method. While some findings have supported the use of inductive methods (Haslerud & Meyers, 1958; Kersh, 1958; Ray, 1961), others favor the use of deductive methods (Forgus & Schwartz, 1957; Sobel, 1956), while still other research has found the two methods to be about equally effective (Craig, 1956; Fowler, 1931). However, it is conceivable that both of these methods might be highly effective but for different subgroups of learners, hence requiring study of interactions.

While results from among the few experiments exploring this general area (Becker, 1967; Kropp, Nelson, & Ring, 1967; Tanner, 1968) have been either negative or highly ambiguous, a survey of literature (Cronbach & Snow, 1969) provides a number of indications that learner characteristics may influence the extent to which learning from inductive and deductive methods occurs. Thus, the purpose of this experiment was to investigate the effects of individual differences on learning from inductive and deductive instructional procedures. An additional variable of interest concerned the frequency with which rules and examples were alternated in inductive and deductive instructional sequences. It was anticipated that ability–performance relationships might conceivably be systematically altered through manipulation of this instructional variable. The medium of programmed instruction was used to permit holding instruction constant in addition to controlling the amount of learner activity and the amount of feedback, which were the same for both inductive and deductive sequences.

---

[1] The author wishes to acknowledge the data processing assistance obtained from the Research and Development Center for Teacher Education, University of Texas.

Requests for reprints should be sent to Mary Lou Koran, Department of Educational Psychology, University of Texas, Austin, Texas 78712.

The requirements of the inductive and deductive sequences of programmed instruction were believed to be sufficiently different to produce different ability–performance relationships. Such hypotheses imply nonparallel regression slopes—specifically that one treatment will not be superior throughout the aptitude distributions.

Tentative predictions concerning ability–performance relationships were derived using Melton's (1967) theoretical model for investigating individual differences in learning as a heuristic device in attempting to select and organize task and ability variables. Melton distinguished three components in associative learning: stimulus differentiation, representing the subjects' coding response to the physical stimulus; association, representing direct or mediated linkage between the encoded stimulus and the required response; and response integration representing the required output response or combination of responses. Differences and similarities in task and ability variables involved in the inductive and deductive programmed sequences can be represented in terms of this model.

In the stimulus differentiation component of the model, the subjects in both inductive and deductive programmed sequences read the program in which they encode verbal information and can establish their own pace. These task variables would be expected to require the ability to comprehend written verbal material.

Corresponding to the associational–mediational component of the model, the subjects in the inductive sequences are given a set of problems or examples from which specific rules must be derived and applied in solving subsequent problems. These task characteristics would appear to require ability in formulating and testing hypotheses fitting a given set of data as well as considerable reasoning ability in structuring mathematical problems. While verbal ability is required in both inductive and deductive sequences, it is presumed to be more decisively involved in the inductive sequences. The necessity of abstracting a verbal rule statement rather than adopting the program's carefully chosen rule statement would appear to constitute a heavier verbal demand.

In the deductive sequences, the subjects are provided with a verbal rule statment which must subsequently be used in solving sets of related problems. This would appear to primarily require ability to reason from a stated rule to a correct solution. Although subjects in the deductive sequences must also solve problems requiring mathematical reasoning, the problems have, to a large extent, been prestructured by the rule statment.

Finally, in the response integration component of the model, subjects in both treatments must state the rules and execute the appropriate arithmetic operations. In doing so, the subjects in both sequences must remember the rule statements as well as keep track of previous applications. While memory ability would be expected to be required in both sequences, the previous derivation of the rules in the inductive sequences would presumably have caused learners to become more cognizant of relationships and hence, minimize the memory requirement. Consequently memory would appear to be more decisively involved in the deductive sequences.

Thus while each of the ability variables considered was expected to be involved in both inductive and deductive programmed sequences, the treatment requirements were hypothesized to be proportionally different. The foregoing analysis can be considered only a partial analysis in that it is confined to the learning task, and thus directly related to some criterion variables while only indirectly related to others. For example, transfer of training may be expected to be dependent upon initial learning. However the extent to which ability–performance relationships may be expected to change from learning to transfer tasks cannot readily be determined. Previous research (Cronbach & Snow, 1969) has indicated that such relationships may be exceedingly complex, and the present state of theory and research does not provide a basis for sophisticated prediction.

METHOD

Subjects

The experimental sample consisted of 167 upper-division students at the University of Texas who were enrolled in an introductory course in the teacher education program. All subjects were exposed to the experimental procedures prior to receiving any formal instruction in educational measurement.

Materials

Experimental procedures were based upon programmed materials and concomitant criterion measures developed by Krumboltz and Yabroff (1965). A linear sequence, constructed response, and horizontal format were used for the program. The programmed subject matter consisted of concepts selected from the area of elementary statistics and test interpretation including mean, median, standardization sample, normal curve, and grade equivalent. Each concept in the program contained both rule frames—stating a definition, rule, or general principle, and example frames—providing applications, examples, and problems—for that concept.

Four forms of the program were utilized; two inductive sequences and two deductive sequences. All four booklets contained 117 frames. A total of 33 rule frames and 84 example frames were presented. While there were only nine different rules, each rule was restated three or four times. The frames used in each booklet were the same; only the order of presentation varied. Deductive sequences presented rule frames prior to example frames for a given concept, while inductive sequences presented example frames prior to rule frames. In addition, the order of the rule frames was modified to accommodate differences between inductive and deductive sequences such that in the deductive sequences, for a given concept the number of cues provided by successive rule statements proceeded from maximal to minimal, while in the inductive sequences the order of the rule frames was exactly reversed. Two arrangements of each method were used; low and high frequency of alternation. In the low-alternation sequences, all of the rule frames for a given concept were presented together and all of the example frames were presented together. In the high-alternation sequences, both rule and example frames for a given concept were interspersed providing frequent shifts between rules and examples. Thus a 2 × 2 factorial design was employed consisting of two instructional methods and two alternation frequencies.

Procedure

Ability measures selected from the Kit of Reference Tests for Cognitive Factors (French, Ekstrom, & Price, 1963), included Letter Sets (Induction), Vocabulary (Verbal Comprehension), First and Last Names (Associative Memory), Logical Reasoning (Syllogistic Reasoning), Knowledge of Arithmetic Operations (General Reasoning), and Auditory Number Span (Memory Span). The abilities assessed were believed to be those which distinguished most clearly between the inductive and deductive programmed sequences and were consistent with the analysis of task and ability variables derived.

Following aptitude test administration, the subjects received one of four forms of the programmed booklet; deductive–low alternation, deductive–high alternation, inductive–low alternation, or inductive–high alternation. The four forms of the programmed booklet were arranged in random order and distributed to subjects. Upon completion of the booklet, materials were returned to the experimenter and the time spent during instruction was recorded. To ensure that subjects actually read the booklet rather than simply copying the answer frames onto the answer sheets, there were four places in the booklet in which the subjects were asked to circle frame numbers on their sheet. The subjects failing to circle at least three of the four numbers were eliminated from the experiment.

A delayed criterion test was administered 2 weeks after the completion of the programmed booklets. The test was given as a unit but consisted of two subtests; a Knowledge of Rules subtest comprised of 19 items, in which the subjects were asked to state the rules from cues other than those found in the program itself, and a Problem Solving subtest containing 29 items measuring the ability to solve previously unencountered problems related to concepts taught in the programmed booklet. Two forms of the test were randomly distributed to the subjects; one in which the Knowledge of Rules subtest was presented first and one in which the Problem Solving subtest was presented first. This was done to ensure that learning resulting from the criterion test administration itself was equalized for the two subtests with no systematic advantage accruing for any particular treatment combination.

Dependent Variables

Any instructional treatment has multiple effects, and the effectiveness of instructional methods may vary for a number of different outcomes (Cronbach, 1966). Accordingly, in the present study a number of different criterion measures were used:

1. Errors during instruction: The number of self-reported errors by each subject on program frames.
2. Time during instruction: The number of minutes required to complete the programmed booklet.
3. Criterion test score: The number of correct answers on the Knowledge of Rules subtest and the Problem Solving subtest.
4. Time on criterion test: The number of minutes required to complete the Knowledge of Rules subtest and Problem Solving subtest.

RESULTS

Treatment Effects

A 2 (methods) × 2 (alternation frequencies) analysis of variance was used to test instructional treatment main effects. Means and standard deviations of all measures of the dependent variables and the pooled within-treatments intercorrelations among dependent variables are reported in Tables 1 and 2.

Analysis of variance results failed to attain significance on criterion test scores, criterion test time, and program time. Thus from the average data alone, inductive and deductive instructional procedures, as defined in this experiment, were about equally effective in terms of time required to teach the given body of information and in promoting transfer of training. However, subjects in the inductive treatments made significantly more program errors ($F = 8.48$, $df = 1/163$, $p < .01$) than did subjects in the deductive treatments.

TABLE 2

INTERCORRELATIONS AMONG DEPENDENT VARIABLES

Criterion	1	2	3	4
1. Criterion test score		−01	−06	−26
2. Criterion test time			60	11
3. Program time				11
4. Program errors				

Note.—Decimals points are omitted.

The results obtained in the evaluation of instructional treatment main effects correspond closely to results reported by Krumboltz and Yabroff (1965) using these same programmed materials and criterion measures. These results generally failed to confirm expectations that specific combinations of inductive and deductive sequences of instruction with varying frequencing of alternation between rules and examples would effect learning differentially. It has been suggested (Krumboltz & Yabroff, 1963) that the highly intriguing nature of the puzzle-type tasks used in experiments commonly cited in support of inductive instructional procedures may reflect a specific rather than general advantage of inductive methods. Experiments using more conventional classroom material (Wolfe, 1963) have tended to produce results similar to those reported in the present experiment.

Moreover, the lack of significant differences between high and low frequencies of alternation may suggest that in a short pro-

TABLE 1

MEANS AND STANDARD DEVIATIONS OF DEPENDENT VARIABLES

Criterion	Treatment group							
	Deductive-Low		Deductive-High		Inductive-Low		Inductive-High	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Criterion test score	25.47	7.51	24.74	6.86	25.88	7.46	25.54	5.64
Criterion test time	26.46	5.87	28.02	5.91	26.98	6.89	27.07	7.28
Program time	78.12	15.31	84.90	17.41	80.52	21.52	85.24	23.24
Program errors	9.76	7.08	11.00	7.23	15.20	7.95	12.31	7.49

gram, such as this, sequencing may not be as important as it would be in longer programs where longer periods of instruction may be spent on abstractions before proceeding to examples or vice versa. However, research showing that some programs teach as well when the frames are arranged in random sequence as when the frames are presented in some ordered sequence suggests that students may learn effectively in some instances, regardless of the sequence (Levin & Baker, 1963; Payne, Krathwohl, & Gordon, 1967).

Aptitude × Treatment Interactions

Aptitude × Treatment interactions were evaluated by comparing regression slopes obtained for aptitude-criterion pairs using *F* tests for heterogeneity of regression. These results are summarized in Table 3. In view of the exploratory nature of this research, single rather than multiple variable regression analysis was used, despite possible inflation of probabilities in some instances. All reported interactions were disordinal such that in each case, regression lines intersected within the range of the aptitude variable considered. Such findings imply, in general, that differential assignment of subjects on opposite sides of the intersection to alternate treatments is appropriate for maximizing learning outcomes.

Analyses of Aptitude × Treatment interactions disclosed that Vocabulary test scores interacted significantly with presentation conditions for program time and criterion test time. In each case, overall significance was primarily due to differences between inductive high-alternation and deductive low-alternation treatments with slopes for inductive low-alternation and deductive high-alternation treatments falling between. Vocabulary scores were negatively related to the time variables in the inductive high-alternation treatment while not significantly related to time in the deductive low-alternation treatment.

In addition, scores for Necessary Arithmetic Operations produced significant disordinal interactions with both program time and criterion test time. Again, differences between inductive high-alternation and deductive low-alternation treatments appeared to contribute most heavily to interaction effects. Test scores were negatively related to these time variables in the inductive high-alternation treatment, while not significantly related to time in the deductive low-alternation treatment.

Moreover, scores on Letter Sets produced significant interactions for program time and criterion test time. Test scores were negatively related to program time in deductive high-alternation and inductive low-alternation treatments while not significantly related to program time in deductive low-alternation and inductive high-al-

TABLE 3

SUMMARY OF SIMPLE REGRESSION ANALYSES SIGNIFICANT DISORDINAL INTERACTIONS

Criterion	Predictor		Treatment groups								F
	Aptitude test	SD	Deductive-Low		Deductive-High		Inductive-Low		Inductive-High		
			a	*b*	*a*	*b*	*a*	*b*	*a*	*b*	
Program time	Verbal Comprehension	3.54	78.02	.61	84.32	−1.91**	80.57	−2.09**	85.84	−2.55**	2.97*
Program time	Necessary Arithmetic Operations	2.36	78.72	−1.50	82.51	−4.76**	80.60	−5.44**	86.29	−6.44**	2.64*
Program time	Letter Sets	2.50	77.77	−.24	85.03	−3.93**	80.60	−3.98**	85.37	−.53	3.21*
Criterion test time	Verbal Comprehension	3.54	26.40	.20	27.95	−.28	26.95	−.34	27.23	−.92**	2.60*
Criterion test time	Necessary Arithmetic Operations	2.36	26.48	−.17	27.62	−.32	26.70	−.17	27.25	−1.10**	3.03*
Criterion test time	Letter Sets	2.50	26.53	.07	27.97	−1.29**	26.99	−.64*	27.05	.55	2.61*

* $p < .05$.
** $p < .01$.
Note.—a and b are of the form $\hat{Y} = a + bx$.

ternation treatments. Letter Sets scores interacted in a similar manner with criterion test time. However, while the same general pattern was obtained, in this case, overall significance was primarily due to differences between the deductive high-alternation and inductive low-alternation slopes. The regression slope obtained from the deductive high-alternation treatment was negative, whereas Letter Sets scores were not significantly related to criterion test time in the inductive high-alternation treatment.

DISCUSSION

It will be recalled that the major purpose of this experiment was to examine the effects of aptitude variables on learning from inductive and deductive instructional procedures. A theoretical framework was suggested as a basis for tentative predictions regarding ability–performance relationships. However, it is apparent that relations between aptitude and instructional procedures are much more complicated than appears on the surface. Any treatment has multiple effects, and different criterion measures may be related to different aptitudes. In the present experiment, aptitude measures interacting with treatment conditions for time variables did not distinguish between treatments on criterion test scores or errors during instruction. Thus any differential assignment based on Aptitude × Treatment interactions must consider which specific learning outcomes are to be obtained. Interestingly, while the aptitude measures used failed to interact with inductive and deductive programmed sequences for the actual criterion test scores, they did produce significant disordinal interactions for time variables. There are both cost and efficiency implications in such results, since the cost of training may be considerably reduced as time required to attain various objectives is minimized. It is possible that these findings might be interpreted in support of the contention (Skinner, 1954) that with proper linear programming, individual differences may be expected to primarily effect time variables. However, alteration of the programmed materials and criterion measures might conceivably produce quite different results. Research previously reviewed in this general area (Cronbach & Snow, 1969) has reported results similar to those obtained in this experiment with respect to criterion test scores. However, these experiments have not included time variables. Such findings would appear to emphasize the usefulness of considering a broader range of dependent variables in studies of this type (Cronbach, 1966).

In addition, an inspection of the direction of Aptitude × Treatment interactions obtained suggestions that any explanation of the interactions generally requires the simultaneous consideration of both method and alternation frequency effects. Each of the obtained Aptitude × Treatment interactions involved aptitude variables believed to enter into the mediational component of the model used to select and organize task and ability variables (Melton, 1967), including Verbal Comprehension, Necessary Arithmetic Operations, and Letter Sets scores. These interactions appeared to support initial hypotheses only in a limited sense. For example, it was anticipated that inductive procedures would produce heavier verbal demands in view of the requirement of deriving a verbal rule statement. Similarly, mathematical reasoning requirements were expected to be greater for inductive procedures due to lack of rule information before solving example problems. However, while Verbal Comprehension and Necessary Arithmetic Operations scores did interact with inductive and deductive procedures for program and criterion test time, interactions were primarily due to differences between inductive high-alternation and deductive low-alternation sequences.

In contrast to these findings are the interactions obtained using Letter Sets scores with program and criterion test time, in which test scores were negatively related to time in deductive high-alternation and inductive low-alternation treatments. Such results are difficult to interpret. While it is possible that these findings could be due to chance, this general pattern of results was consistent with respect to the relationship between Letter Sets scores and both time variables. These findings do illustrate, however, that labels representing systematic differences in experimental procedures do not necessarily reflect systematic differences in aptitudes or processes involved within those treatments.

An examination of the general pattern of the interactions obtained shows that in each case, the slope for the deductive low-alternation treatment appeared relatively flat, whereas the other treatments, especially the inductive high-alternation treatment, produced steeper slopes. The deductive low-alternation treatment may conceivably compensate to the greatest extent for deficiencies in verbal processing and mathematical structuring through its explicitly derived rule-example–incomplete-example presentation. With such a sequence, low-ability subjects can recognize and apply a rule with proficiency while to approach a rule through the high-alternation sequence or through induction may be expected to be a slower more laborious process.

Conversely, in the less obvious, explicitly derived sequences such as the inductive high-alternation treatment with its frequent shifting from example to rule derivation, subjects high in verbal and mathematical reasoning appear to do well with respect to the time variables. Possibly the heavier verbal and mathematical mediational requirements constitute a preferred, more challenging mode of operation for these subjects, in which they may rapidly isolate and use key stimuli in frequent shifts from example to rule derivation. In contrast, subjects low in verbal and mathematical reasoning are not compensated by this treatment. Thus, Aptitude × Treatment interactions may be viewed as a compensatory-conciliatory process in which some harmony is achieved between task and learner characteristics (Cronbach & Snow, 1969; Koran, Snow, & McDonald, 1971).

Experiments such as these have only begun to explore the wide range of problems concerned with finding effective instructional techniques for students with different characteristics. Additional task analysis may indicate a variety of ways in which aptitude measures, programmed materials and criterion measures may be varied to alter the results obtained. Certainly further investigation should precede attaching practical significance to these results. This experiment is useful in indicating how complex such experimentation is likely to be. As Cronbach and Gleser (1965) have suggested, initial study of Aptitude × Treatment interactions will quite possibly be more important for what it tells us about the psychology of instruction than for immediate placement purposes. However, such experimentation may eventually help provide a basis for the individualization of instruction.

REFERENCES

BECKER, J. P. An attempt to design instructional techniques in mathematics to accommodate different patterns of mental ability. Unpublished doctoral dissertation, Stanford University, 1967.

CRAIG, R. C. Directed versus independent discovery of established relations. *Journal of Educational Psychology,* 1956, **47,** 223–234.

CRONBACH, L. J. The two disciplines of scientific psychology. *American Psychologist,* 1957, **12,** 671–684.

CRONBACH, L. J. The logic of experiments on discovery. In L. Shulman & E. Keislar (Eds.), *Learning by discovery: A critical appraisal.* Chicago: Rand McNally, 1966.

CRONBACH, L. J. How can instruction be adapted to individual differences. In R. Gagné (Ed.), *Learning and individual differences.* Columbus, Ohio: Charles E. Merrill, 1967.

CRONBACH, L. J., & GLESER, G. C. *Psychological tests and personnel decisions.* (2nd ed.) Urbana: University of Illinois Press, 1965.

CRONBACH, L. J., & SNOW, R. E. Individual differences in learning ability as a function of instructional variables. (Final Report, OEG 4-6-061269-1216) Washington, D.C. United States Government Printing Office, 1969.

FORGUS, R. H., & SCHWARTZ, R. J. Efficient retention and transfer as affected by learning method. *Journal of Psychology,* 1957, **43,** 135–139.

FRENCH, J. W., EKSTROM, R. B., & PRICE, L. A. *Kit of Reference Tests for Cognitive Factors.* Princeton, N.J.: Educational Testing Service, 1963.

FOWLER, H. I. *Induction or deduction? An experimental investigation in the psychology of teaching.* (Educational Research Series, No. 5, Australian Council for Educational Research) Melbourne: Melbourne University Press, 1931.

HASLERUD, G. M., & MEYERS, S. The transfer value of given and individually derived principles

Journal of Educational Psychology, 1958, **49,** 293–298.

KERSH, B. Y. The adequacy of "meaning" as an explanation of learning by independent discovery. *Journal of Educational Psychology,* 1958, **49,** 282–292.

KORAN, M. L., SNOW, R. E., & MCDONALD, F. J. Teacher aptitude and observational learning of a teaching skill. *Journal of Educational Psychology,* 1971, **62,** 219–228.

KROPP, R. P., NELSON, W. H., & RING, F. J. Identification and definition of subject-matter content variables related to human aptitudes. (Cooperative Research Project No. 2914) Unpublished report, Florida State University, 1967.

KRUMBOLTZ, J. D., & YABROFF, W. The comparative effects of inductive and deductive sequences in programmed instruction. *American Educational Research Journal,* 1965, **2,** 223–235.

LEVIN, G. A., & BAKER, B. L. Item scrambling in a self instructional program. *Journal of Educational Psychology,* 1963, **54,** 138–143.

MELTON, A. Individual differences and theoretical process variables: General comments on the conference. In R. Gagné (Ed.), *Learning and individual differences.* Columbus, Ohio: Charles E. Merrill, 1967.

PAYNE, D. A., KRATHWOHL, D. R., & GORDON, J. The effect of sequence on programed instruction. *American Educational Research Journal,* 1967, **4,** 125–132.

RAY, W. E. Pupil discovery vs. direct instruction. *Journal of Experimental Education,* 1961, **29,** 271–280.

SKINNER, B. F. The science of learning and the art of teaching. *Harvard Educational Review,* 1954, **25,** 86–97.

SOBEL, M. A. Concept learning in algebra. *Mathematics Teacher,* 1956, **49,** 425–430.

TANNER, R. T. Expository-deductive vs. discovery-inductive programing of physical science principles. Unpublished doctoral dissertation, Stanford University, 1968.

WOLFE, M. S. Effects of expository instruction of mathematics on students accustomed to discovery methods. Unpublished doctoral dissertation, University of Illinois, 1963.

(Received March 31, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 4, 308–314

EFFECTS OF ENCOUNTER WITH VERBAL UNITS IN PROSE CONTEXTS ON SUBSEQUENT ASSOCIATIVE LEARNING

ROGER H. BRUNING[1] AND LARRY J. LANTINGA[2]
University of Nebraska

Unfamiliar bisyllables appeared as concept exemplars in descriptive prose materials as either different exemplars of one concept (convergent materials) or each representing a different concept (nonconvergent materials). Extrapolating from studies of paired-associate intralist similarity, it was predicted that study of convergent materials would interfere with subsequent paired-associate learning involving the units. Where units became paired-associate responses, pretraining conditions pairing familiarization with irrelevant prose, and with nonconvergent prose, required significantly fewer trials to criterion than the familiarization–convergent condition ($p < .01$, $< .05$). The familiarization–convergent condition was no better than a control condition, which involved no prior contact with the units. Findings were nonsignificant for units as paired-associate stimuli. Implications are that shared associates to units acquired from prose interfere with associative learning when units are used as responses, with effects depending on the unit role in subsequent learning.

Written materials, particularly those with instructional intent, frequently involve an arrangement in which previously unfamiliar verbal units are presented to the learner in some associational framework, often as exemplars for conceptual categories. An individual's learning history may have included reading at one time or another, for example, that the plover is a bird, that tin is an element, that Paris is a city, and so forth. Ordinarily, additional descriptive information is presented at the same time with the intention that the learner might also attach this information associatively to the previously unfamiliar units. Thus, we may also have learned in the same passages that the plover has long legs and pointed wings, that tin is metallic and malleable, and that Paris is located on the Seine.

The present study explored two questions which are conditional on the assumption that associations such as these can be and are acquired from written materials. The first such conditional question was the following: Will associations to verbal units presented in written materials have effects on subsequent learning involving the units? In order to deal with the question adequately, the interrelatedness of frequency of encounter with verbal units and increases in number of associations must, of course, be recognized. It is clear that any attempt to establish or increase associates through learner interaction with verbal units in written materials is simultaneously contributing to familiarization with the unit itself in terms of simple frequency of encounter. Prior research indicates that an increase along either dimension, frequency or associations, tends to result in facilitation of learning. With regard to frequency of encounter, Underwood and Shulz (1960), for example, found that facilitation of learning was produced by response familiarization in a paired-associate task, although stimulus familiarization did not produce that effect. In an investigation relevant to the present question of effects of associations, Cieutat, Stockwell, and Noble (1958), in a representative study, found that increases in rated meaningfulness of units, a measure of number of associations, were related to improvement in paired-associate learning. Effects were somewhat greater for response than for stimulus meaningfulness.

The present study contained two pretraining manipulations involving verbal units. First, under what were called familiarization conditions, subjects studied a list of eight low-meaningful verbal units for a short period of time. A second type of manipulation involved encounter with the units in prose contexts in which the units were presented as names or exemplars for concepts, which offered opportunity for both further familiarization and associative learning. Major treatment conditions involved combinations of these two types of manipulations. Criterion performance was assessed on a paired-associate task in which units were used either as stimuli or responses. It was predicted that learning in the paired-associate task would be facilitated over control conditions for any conditions which involved familiarization. For those conditions which involved both familiarization and encounter of units in written associative contexts, further facilitation of subsequent paired-associate learning was predicted. Effects of all conditions were predicted to be somewhat stronger when units were employed as paired-associate responses than when they were used as paired-associate stimuli.

---

[1] Requests for reprints should be sent to Roger Bruning, 15 Teachers College, University of Nebraska, Lincoln, Nebraska 68508.

[2] Appreciation is extended to Kenneth D. Orton and Royce R. Ronning for their critical comments.

Of considerable interest was a second conditional question regarding conceptual relatedness of associations. That is, if units in the written materials shared a common "conceptual" associate, what effect might this have upon the paired-associate learning? Results of earlier experimentation indicate that while similarity of responses may result in facilitation of response learning in a paired-associate task (Underwood, Runquist, & Schulz, 1959) or at least produce no effect on response learning (Jung, 1965), response similarity may produce interference in the associative stage of paired-associate learning (Horowitz, 1962; Jung, 1965). While the findings of the latter studies were with regard to similarity along formal dimensions for CCC trigrams and CVC nonsense syllables, interference in associative learning due to response similarity is also suggested in the data of Underwood, Runquist, and Schulz, who investigated effects of similarity along a dimension of meaning, although associative learning was not under direct investigation in their study. In one major treatment condition of the present study, each of the unfamiliar verbal units was encountered in prose materials as an exemplar for a common conceptual category; that is, each was presented as a different representative of a single concept class. In a contrasting condition, the units were encountered in similar materials but each as an exemplar of a *different* concept class. It was predicted that subjects in the first condition would have more difficulty in the paired-associate task than would subjects in the second. Although the total amount of attributive information was held constant across these conditions, it was reasoned that the formation of nonconvergent associations to the units in the latter case would be more compatible with the requirements of the paired-associate task. On the other hand, associations mediated through a single conceptual category, called convergent associations, were expected to create interference in the associative stage of paired-associate learning and, analogous to the effects which the earlier researchers had found with regard to response similarity, produce a relative decrement in performance. Again, effects were predicted to be more pronounced when units were used as paired-associate responses than when they were used as paired-associate stimuli.

METHOD

Subjects

The subjects were 64 students from an upper division undergraduate educational psychology course. Participation in the experiment was a course requirement.

Materials

Sixty-four paragraphs of an expository nature were prepared as study materials for use in the initial phase of study. Eight paragraphs described exemplars for each of eight conceptual categories: birds, deserts, fish, islands, lakes, peoples (tribes), plants, and rivers. Each paragraph contained four major elements of attributive factual information relative to the exemplar in addition to stating the concept class to which the exemplar belonged. One class of attribute, location or locale where the exemplar could be found, was stated for each exemplar while the other three attributes were unique in the sense that they could not be subsumed within any one class of attributes. Each paragraph describing an exemplar consisted of three sentences incorporating the four attributes and presented the name of the particular exemplar twice, once in the first and once in the last sentence of the paragraph.

As exemplars for the concept classes, eight words from among the 24 lowest ranking bisyllables in terms of association value were selected from Cieutat's (1963) list of randomly selected monosyllables, bisyllables, and trisyllables from the English language. Words selected were the following: *cuadra, elench, gossan, illeck, markaz, pagrus, santon,* and *tabret*. Criteria of acceptability other than that they were low-meaningful and bisyllabic included plausibility as "names" and independence from one another in terms of pronounceability. Judgments of acceptability on these latter criteria were made jointly by the experimenters. Association values ranged from .46 to .58. Also, for joint use with the "names" in a subsequent paired-associate learning task, eight monosyllables of four letters with the highest association values for that group were selected from the Cieutat list. Words chosen were *suit, make, arms, well, buck, drum, dish,* and *tide*. Association values were 1.00 for all, with the exception of a .99 rating for *make*.

Procedure

The study consisted of two phases: (*a*) a prose materials study pretraining session of 20-minute duration followed by a short 10-minute quiz over the materials and (*b*) a paired-associate learning task. In the first phase, subjects were taken individually and assigned at random as they arrived at the laboratory to one of four study conditions: familiarization and nonconvergent written materials, familiarization and convergent written materials, familiarization and irrelevant materials, and irrelevant prose materials pretraining. Study materials for all conditions were contained in a booklet. Instructions printed on the front stated that the booklet contained material for study, that instructions printed inside were to be followed, that subjects were to proceed one page at a time, and that the study period for all of the materials would be limited to 20 minutes.

In conditions involving familiarization, subjects were first presented with a list of the eight bisyllables as "foreign names" to be learned. Instructions were as follows: "Write each of the following foreign names three times in the spaces provided and study them carefully until you can write them from memory. The order in which you learn them is not important." Order of the lists was randomized across subjects. At the bottom of the page were instructions to continue to the next page and to limit study on the current page to no more than 5 minutes. On the following page were eight blanks and instructions to then check their words and correct any errors. Again a further note advised them to spend no more than 5 minutes on that page. The written prose paragraphs followed on the next page, prefaced by the instruction: "The following is a sample of information taken from a world almanac. Study over it carefully in preparation for a posttest which will cover a number of the facts presented."

For the familiarization–nonconvergent written materials condition, the study paragraphs were formed as follows, with three variables entering into the final formulation of subjects' written materials: (*a*) the concept to be represented, (*b*) the particular paragraph from among the set of eight describing exemplars of a concept, and (*c*) the "name" which would be assigned into the paragraph. Thus the selection of the concept, the choice of paragraph describing an exemplar, and the names assigned to it were accomplished by randomly drawing, without replacement, numbers from three sets of eight numbers. A number drawn from the first set determined the concept class from which the paragraph would come (e.g., rivers), the number from the second determined which descriptive paragraph would be chosen (e.g., "______ is a river located in Central America . . ., etc."), and the third number determined which of the eight names would be assigned into the paragraph (e.g., Gossan). The materials for each subject in the familiarization–nonconvergent condition thus consisted of paragraphs selected in the above manner, one paragraph from *each of* the eight concept classes to which a name had been randomly assigned. Thus, the written materials for a subject may have presented Gossan as a river with certain characteristics, may have described Markaz as a plant with certain attributes, and so on for this subject within the familiarization–nonconvergent condition while for another subject Gossan may have been encountered as a name for a bird or a fish. The 64 paragraphs were used without replication over each group of eight subjects in the familiarization–nonconvergent condition.

For the familiarization–convergent condition, the same 64 paragraphs as were used in the familiarization–nonconvergent condition were also encountered across each group of eight subjects, but an individual's study materials consisted of eight paragraphs selected *within* a particular concept category, rather than one from each as in the familiarization–nonconvergent condition. Again, the same names were randomly assigned to paragraphs for each subject. Thus, the study materials for one subject in this condition might have consisted of eight paragraphs about birds, named, of course, randomly as before, while another studied about eight rivers, and another about eight deserts. A sample paragraph from among the 64 paragraph pool of materials for the familiarization–convergent and familiarization–nonconvergent conditions which reflects a particular verbal unit-paragraph combination is presented below:

> Cuadra is an island in the South Atlantic Ocean, located about 1200 miles west of Africa. It has a pleasant climate with an average annual rainfall of about 40 inches. The inhabitants of Cuadra are of mixed African origins.

In the familiarization-irrelevant materials condition, students participated in the reading, writing, and recall of the bisyllables in the familiarization phase, but studied an irrelevant passage of factually oriented material on desalinization of sea water of comparable length to the eight paragraphs. In the irrelevant prose materials pretraining condition, subjects read only the passage on desalinization and had no contact with the experimental bisyllables prior to the paired-associate task.

At the end of a total elapsed time of 20 minutes from beginning work on the study booklet, the examiner picked up the booklet and administered an eight-question short-answer posttest over the materials. Posttests for subjects studying paragraphs in the familiarization–convergent and the familiarization–nonconvergent conditions were constructed such that each item sampled information from a separate paragraph and that items alternately quizzed (*a*) information regarding where the exemplar studied could be found and (*b*) a distinguishing characteristic of a particular exemplar. Thus, questions were parallel for these groups in the sense of sampling the same classes of attributes, although they were, at the same time, idiosyncratic to the particular materials studied by a given subject. The posttest for the familiarization–irrelevant materials and the irrelevant prose materials pretraining conditions consisted of eight short-answer recall items sampling information from the desalinization passage.

Upon completion of the approximately 30-minute study and quiz session, subjects were directed individually by the first experimenter to a separate experimental room for the paired-associate task. The nature of a paired-associate task was explained by the second experimenter and a warm-up task consisting of four pairs of high-frequency bisyllables was administered. In the practice task, pairs were learned to a criterion of two consecutive correct trials. Following the practice session, any questions were answered. In this warm-up task, as in the subsequent experimental paired-associate task, the paired-associate task was administered by Experimenter 2 who sat across a small table facing the subject. Paired-associate materials were contained in a notebook with one pair printed per page. Administration of the paired-associate task was by the study trials–test trials method at a 5-second presentation and recall interval. On study trials subjects were required to pronounce both words as a pair appeared. For test trials, the first word was printed, followed by a blank and a question mark, and the subject was required to pronounce the stimulus word and to write the correct response in the blank. Experimenter 2 turned each page in the materials for the subject, timing the turns by means of an auditory signal occurring at 5-second intervals and delivered to Experimenter 2 by earphone from a prerecorded tape. The sound was not audible to the subject. Subjects were assigned to one of two paired-associate conditions: (*a*) response, in which the eight bisyllables encountered in the pretraining (except for subjects in the irrelevant prose materials pretraining condition) served as responses while the eight high-frequency monosyllables served as stimuli and (*b*) stimulus, in which the bisyllables served as stimuli and the monosyllables were responses. Twelve different random pairings of the eight stimuli with the eight response terms were employed in nonsystematic fashion for the 32 subjects in the response condition and another 12 pairings for the 32 in the stimulus condition. Because of the obvious potential for experimenter bias, the two phases of the experiment were conducted independently, with Experimenter 2 having no knowledge of the pretraining treatment for any subject. In the paired-associate task, subjects were run sequentially, four under the stimulus conditions and four under the response conditions. At the same time, Experimenter 1 followed the strategy of randomly assigning booklets of the four pretraining conditions within each group of four successive subjects. Thus, each successive group of eight subjects had one entry into each of the eight possible pretraining paired-associate combinations: familiarization–nonconvergent written materials–stimulus, familiarization–convergent written materials–stimulus, familiarization–irrelevant materials–stimulus, irrelevant prose materials pretraining–stimulus, familiarization–nonconvergent written materials–response, familiarization–convergent written materials–response, familiarization–irrelevant materials–response, irrelevant prose materials pretraining–response.

RESULTS

Mean trials to criterion for stimulus and response forms of the paired-associate task are presented in Table 1, while paired-associate learning curves for the response pretraining conditions are presented in Figure 1. Since stimulus and response conditions did not seem to be directly comparable because of the obvious differences between familiar monosyllables and unfamiliar bisyllables as paired-associate responses, stimulus and response data were analyzed separately by means of independent one-way analyses of variance. When initial inspection of the response data indicated the possibility of nonhomogeneity of cell variances a test of this factor was made using Levene's technique, as suggested by Glass (1966). However, the degree of heterogeneity was not significant ($F = 2.52$, $df = 3/28$, $p > .05$) so no transformations of the raw scores were made. Overall differences among the four pretraining conditions were highly significant for the response data ($F = 14.71$, $df = 3/28$, $p < .001$), while differences for the stimulus condition approached significance ($F = 2.35$, $df = 3/28$, $.10 > p > .05$). Further analyses of differences between means in the response data using Scheffé contrasts as conservative tests of significance, revealed that the familiarization–irrelevant materials condition required

TABLE 1

NUMBER OF TRIALS TO CRITERION ON A PAIRED-ASSOCIATE (PA) TASK FOR GROUPS RECEIVING VARYING STIMULUS OR RESPONSE PRETRAINING

	Pretraining condition							
Use of unit in PA task	Familiarization–convergent written materials		Familiarization–nonconvergent written materials		Familiarization–irrelevant materials		Irrelevant prose materials pretraining	
	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
Stimulus	5.25	1.49	7.63	1.99	8.88	4.49	6.75	2.32
Response	12.63[a]	6.41	6.88	3.31	5.63	2.20	11.25	5.44

[a] Paired-associate learning was terminated after 20 trials for two subjects in the response–famaliarization–convergent written materials condition who had not reached criterion. Scores of 20 were assigned to these subjects.

significantly fewer trials to criterion than both the familiarization–convergent ($p < .01$) and irrelevant prose materials pretraining groups ($p < .05$), while the familiarization–nonconvergent condition differed significantly ($p < .05$) from the familiarization–convergent condition in the same direction.

In contrast to findings for the response groups, means for the stimulus conditions, although differences only approached significance ($.10 > p > .05$), showed an order exactly reversing that for the response condition. Whereas subjects under the response conditions had performed in the ascending order—familiarization–irrelevant materials, familiarization–nonconvergent written materials, irrelevant prose materials pretraining, and familiarization–convergent written materials conditions—in terms of number of trials required to criterion, in the stimulus condition, order of means was exactly reversed.

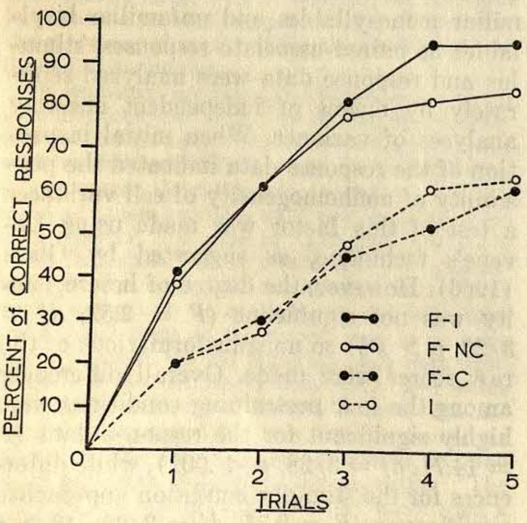


FIG. 1. Percentage of correct responses on a paired-associate task for response pretraining involving familiarization plus irrelevant written materials (F-I), familiarization plus nonconvergent written materials (F-NC), familiarization plus convergent written materials (F-C), and irrelevant written materials only (I).

DISCUSSION

There would appear to be a number of noteworthy aspects to the present data. First, these data support the basic contention that associations to verbal units are acquired from written materials and, more importantly, that the associations have substantial effects on subsequent learning involving the units. These effects, it can be seen, are not necessarily always positive ones, however. Of major interest are those effects which were observed when verbal units which were encountered in varying prose contexts were used as responses in the paired-associate task. It will be recalled that it was predicted that response familiarization plus the encounter with the units in the nonconvergent prose passage would produce greatest facilitation in paired-associate performance, under the expectation that such encounter would produce positive effects along both frequency and associative dimensions. Familiarization alone, it was predicted, would facilitate performance over control conditions but lacking associative learning with regard to the units, facilitation was not predicted to be as great as when it was combined with associative learnings.

However, it can be seen that paired-associate learning was most facilitated by familiarization training alone and that the familiarization plus interaction with prose materials in which units were presented as exemplars of separate conceptual categories together with attributive information was no better than familiarization alone. In contrast to the relatively good performance of the familiarization–irrelevant materials and familarization–nonconvergent written materials conditions, the same familiarization followed by learner study of prose materials in which the units were linked by being presented as exemplars of a common conceptual category (familiarization–convergent written materials) was seen to produce highly inferior performance, performance which was no better than that of the control group (irrelevant prose materials pretraining) which had no prior contact with the units at all. On the basis of these findings it can be argued that the major facilitative function when units were used as paired-associate responses was produced by simple frequency or response familiarization. At the same time, the newly acquired associative responses learned from the written materials appear not to have resulted in improvement in subsequent paired-associate performance, and in the case where units shared a common associate, substantial negative effects were noted. In this latter condition in which units were linked through a common conceptual mediator, the decrement was great enough to negate any effects of the response familiarization.

The results are consistent with earlier findings (Horowitz, 1962; Jung, 1965) which indicated that response similarity may have detrimental effects upon associative learning. The present results differ from, and extend, the earlier findings in that the experimental manipulation here involved the characteristics of prose materials in which units appear, with the materials apparently differentially affecting the probability of elicitation of a common mediating response. In contrast, in the earlier studies the manipulation involved the *selection* of units with regard to some preexisting dimension of similarity.

Although differences for the stimulus conditions only approached significance, ($.10 > p > .05$), the marked variance of these findings from those under the response conditions merits comment. The order of the means was familiarization–convergent written materials, irrelevant prose materials pretraining, familiarization–nonconvergent written materials, and familiarization–irrelevant materials in ascending order of learning difficulty, exactly reversing the order of difficulty which was obtained for the response condition. The implication of this reversal of order would appear to be that those factors producing facilitation in the case of use of verbal units as responses in the present study may have tended to produce decrements in performance when the units were used as stimuli. Additionally, it can be argued, the roles of the units as stimuli are very different than their function as responses. As has been pointed out by Underwood and Schulz (1960), the effective stimulus in a paired-associate task is not necessarily the nominal unit but often some component of the unit. In results reported by these authors (pp. 296–300) relative to learning nonsense syllable pairs, where stimulus and response were linked by associations, most often the associate used was a response to only a single letter of the stimulus, often the initial letter. They argue that if the stimulus is not integrated or does not suggest an immediate association, the subject may merely select a particular part of the stimulus upon which to base his association.

It is suggested that under the stimulus condition the familiarization–nonconvergent written materials and familiarization–irrelevant materials conditions which had the effect, it was argued earlier, of producing well-learned and integrated units, may have been dysfunctional for the paired-associate learning for this reason. It may be that the most effective cues and mediating responses are those which are simplest and convenient for the learner to use, in that they already exist in strength in the habit repertoire of the learner. Pretraining which would tend to cause the learner to now attend to the entire stimulus as a unit and, in the case of the familiarization–nonconvergent condition, to produce some associative responses along newly established meaning dimensions would, perhaps, result in less effective paired-associate learning. Conversely, in the irrelevant prose materials pretraining condition in which subjects had no contact with the units prior to the paired-associate task and units were not integrated, subjects may have been more likely to make the associative hookup in terms of alphabetical components of the stimulus and in terms of stable associative mediators which existed in his repertoire. For the familiarization–convergent condition, it was argued earlier that effects produced through unit familiarization in terms of integration and unit discriminability may have been counteracted through effects of written materials in which discriminability of units was reduced through associative commonality. Under such conditions subjects may have tended to disregard pretraining experience as irrelevant or confusing and operated much as subjects in the irrelevant prose materials pretraining condition, in terms of simpler cues and preexisting associative responses.

To the extent that conditions in this study parallel those generally involved in most verbal acquisition processes from written materials, the present results suggest a need for further exploration of what appear to be quite complex relationships between conditions under which verbal units are initially encountered in prose and subsequent learning or performance involving these units. The great differences between treatment effects which obtained when verbal units were used as stimuli as opposed to their use as responses and differential effects observed due to varying characteristics of associations to units through changes in prose contexts show that no one generalization is appropriate to predict effects of such experiences. Associative learning, for example, was seen to have effects greatly dependent on the degree of interrelatedness of associations to verbal units in prose, to the extent that substantial interference resulted under conditions in which associative relatedness was high. Familiarization with units, as another example, appeared to have substantially different effects depending on the role of units in subsequent learning, producing facilitation when response familiarization was involved but having no positive effect when familiarization was with regard to paired-associate stimuli. To move toward a more complete understanding of these relationships will obviously require further research, although the present data would appear to have begun to delimit some of the relevant variables involved in such types of verbal learning.

REFERENCES

CIEUTAT, V. J. Association indices for 446 randomly selected English monosyllables, bisyllables, and trisyllables. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 176–185.

CIEUTAT, V. J., STOCKWELL, R. E., & NOBLE, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (m, m′) in paired-associate learning. *Journal of Experimental Psychology*, 1958, **56**, 193–202.

GLASS, G. V. Testing homogeneity of variances. *American Educational Research Journal*, 1966, **3**, 187–190.

HOROWITZ, L. M. Associative matching and intralist similarity. *Psychological Reports*, 1962, **10**, 751–757.

JUNG, J. Two stages of paired-associate learning as a function of intralist response similarity (IRS) and response meaningfulness (M). *Journal of Experimental Psychology*, 1965, **70**, 371–378.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *Journal of Experimental Psychology*, 1959, **58**, 70–78.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. New York: J. B. Lippincott, 1960.

(Received April 2, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 4, 315-321

THEMATIC PROMPTING IN PAIRED-ASSOCIATE LEARNING[1]

RICHARD C. ANDERSON,[2] JAMES M. ROYER, RAYMOND W. KULHAVY
SYLVIA D. THORNBURG, AND LAURA L. KLEMT
University of Illinois

The hypothesis that thematic prompts facilitate learning by furnishing mediators was investigated in three experiments. In Experiment I, paired associates were learned as readily from unprompted sentences and unadorned pairs as from thematically prompted sentences displayed as a whole. In Experiments II and III, the prompted sentences were initially exposed with blanks in place of the response terms in order to induce complete processing of the sentences. Thematic prompts enhanced learning under these conditions, perhaps because the prompts provided mediators or maybe because the prompts facilitated encoding of the cues. The effectiveness of a prompt did not depend upon the extent to which it determined the response, a fact which is difficult to understand within standard formulations of mediation.

In the jargon of programed instruction, a *prompt* is a hint included within a frame of a program to help the student answer correctly. A *thematic prompt* makes the correct response more probable by indicating its meaning. The following frame, adapted from Markle's (1969, pp. 66–69) book on programing, illustrates a thematic prompt.

> Pegasus was a most unusual animal. Like a horse, he had four ______, and, like a bird, he had two ______.

The ostensible function of the thematic prompts "like a horse" and "like a bird" is to get the student to write "legs" and "wings." However, this function alone is trivial, not worth further analysis. The frame is interesting only if students learn from it that Pegasus had four legs and two wings, and for all we know that may be accomplished as readily simply by telling them. There are no data which show that thematic prompts facilitate learning or remembering. Indeed, despite the fact that it is a much discussed feature of programmed instruction, thematic prompting has never been experimentally investigated.

Along with whatever other function it serves, a thematic prompt may provide the student with a mediator. There is considerable evidence that mediation is the rule rather than the exception in associative learning. Montague and his associates (Kiess & Montague, 1965; Montague, Adams, & Kiess, 1966; Montague & Wearing, 1967) have shown that the learning of a pair of verbal units is almost always accompanied by the creation of a meaningful connective which they call a "natural language mediator." Under some conditions, connectives provided by the experimenter facilitate associative learning. For instance, Rohwer (1967) has found that children learn to associate pairs of common nouns linked with a verb (The cow chased the ball) or a preposition (The cow behind the ball) more quickly than unlinked pairs. If simply placing verbs or prepositions between nouns improves learning, it is entirely reasonable to believe that the elaborate and distinctive connectives embodied in thematic prompts would be even more facilitative. Consider the following pair of verbal units: SUS-*holes*. Augmenting this pair in the following manner should make it easier to learn:

> After many months of walking, SUS develop *holes*.

Sentences provided by the experimenter do not always facilitate the paired-associate learning of college undergraduates (Kulhavy, 1970). Furthermore, undergraduates learn pairs of words more quickly if they are required to generate sentences linking

---

[1] The research described herein was supported in part by the Advanced Research Projects Agency through the Office of Naval Research under Contract ONR Nonr 3985(08). The authors are grateful for the assistance of Janet Hidde.

[2] Requests for reprints should be sent to Richard C. Anderson, Training Research Laboratory, University of Illinois, 8 Lincoln Hall, Urbana, Illinois 68101.

the pairs rather than to read sentences that the experimenter has prepared (Bobrow & Bower, 1969). One interpretation of this finding is that idiosyncratic sentences are better because each subject selects mediators which have the highest associative strength for him. Experimenter-provided mediators, or thematic prompts, might nonetheless prove superior to subject-generated ones if (*a*) the pair members were difficult to associate and (*b*) the cue and prompt strongly evoked the response.

With respect to the first condition, the reasoning is that when a pair is difficult to associate a thematic prompt will provide a mediator that the subject may not be able to think of himself. On the other hand, when a pair is easy to associate, a thematic prompt may interfere with the subject's own mediator. For instance, the pair, BAK-*guards,* is highly associable. Embedding this pair in the sentence, "To keep people from stealing the money, BAK have *guards,*" may interfere with the performance of people who tend to think of BAKS and *guards* as football players.

It stands to reason that a prompt will not facilitate learning unless it helps to associate the cue and the response. This characteristic can be called "response determination" and it can be operationalized as the probability that the specified response will occur given the prompt and the cue.

The purpose of the research described herein was to explore the effects of thematic prompts on paired-associate learning as a function of cue-response associability and response determination.

EXPERIMENT I

Method

Materials. Thematically prompted sentences were created for 82 CVC-word pairs. These materials were normed in three different ways. First, a group of about 30 undergraduates was given a mimeographed list of the sentences with instructions to supply the English words which would most reasonably replace the CVCs. A sentence was dropped from further consideration unless 90% or more of the subjects gave the encoding of the CVC that the authors had in mind. Second, another group of about 30 undergraduates was given a mimeographed list of the sentences with the appropriate English word in place of each CVC and a blank in place of the response term. The subjects were asked to supply the response terms. The frequency of the intended response terms was the measure of response determination. Finally, still another group of about 30 subjects was given the CVC-word pairs alone with instructions to rate the associability (that is, the ease with which they were able to think of an association) of each pair on a 9-point scale. In unpublished research it has been found that associability ratings correlate very highly with paired-associate learning (see also Montague & Kiess, 1968).

Based on the normative data, a list of 36 sentences was chosen. Moving from left to right, each sentence consisted of the thematic prompt, the CVC, a verb, and the response term. The verb was *is, are, have,* or *has* in 21 of the sentences. Sixteen of the response terms were nouns; the remainder were adjectives. The CVC and the response term in any pair did not possess the same initial letter. The list contained nine hard-to-associate, response-determined items (HD), nine hard, undetermined items (HV), nine easy, determined items (ED), and nine easy, undetermined items (EU). Following are examples of the four types of items:

> In order to keep a belt in its proper place, dungarees and JEN have *loops.* (HD)
>
> After the tanning process is completed, LEH is *tough.* (HU)
>
> Like other vehicles, to cover the engine, TUK have *hoods.* (ED)
>
> For a devout person, BEF are *firm.* (EU).

The items were chosen so as to maximize differences between the sublists in associability and response determination while at the same time equating the mean Thorndike-Lorge (1944) frequency of the response terms. The norming group supplied the correct response 83.1% of the time in determined items but only 9.2% of the time in undetermined items. Moreover, the most frequent response to undetermined items appeared in an average of only 35.3% of the cases and an average of 6.2 different responses per item had to be considered to account for 83.1% of all responses, so the responses really were undetermined in these items. The easy and hard items received average ratings of 3.3 and 7.6, respectively, on a scale where 1 meant very easy to associate and 9 meant very hard to associate.

Subjects, design, and procedure. The subjects were 90 college sophomores enrolled in an educational psychology course, who participated in the experiment to fulfill a course requirement. Each was randomly assigned to one of three conditions. The subject completed study-test trials with either thematically prompted sentences, unprompted sentences (e.g., SUS develop *holes*), or CVC-word pairs. Both study and test trials were presented at a 4-second rate on a Lafayette Model No. 303A memory drum. No overt response was required during study trials. On test trials the CVCs alone were presented; the subject responded orally. There were four randomizations of study materials and four randomizations of the test stimuli to prevent serial learning. Subjects were run to a criterion of 36/36 or 10 trials.

Results

There were no significant differences among conditions overall, on any trial, or any sublist. In fact, the consistent trend was for the paired-associate condition to do slightly better than the unprompted sentence condition, which in turn did slightly better than the thematically prompted sentence condition. There was, of course, substantial improvement over trials, and associability was a significant factor, $F = 83.8$, $df = 1/87$, $p < .01$, which simply confirmed the norming procedures used to classify items as easy or hard. The Associability × Response Determination interaction was also significant, $F = 13.30$, $df = 1/87$, $p < .01$. There was a greater difference between easy- and hard-to-associate pairs when response determination was low than when it was high. No other significant effects were obtained.

In postexperiment interviews, many subjects reported a distressing tendency to ignore the carefully contrived thematic prompts. They said that after one reading they paid little attention to the prompts and tried instead to construct their own mediators. Some stated specifically that the thematically prompted sentences were of no help.

EXPERIMENT II

Among the several conceivable reasons for the failure of thematic prompts to facilitate learning in Experiment I is the possibility that subjects did not use the prompts. A case can be made for the proposition that subjects ordinarily do no more processing of verbal materials than the task demands (Anderson, 1970). Sometimes they may completely ignore aspects of a compound stimulus (cf. Underwood, 1963). But often the processing failure is more subtle. Probably the first thing a person does with words he notices is to say them to himself, that is, to render them into implicit speech. However, there is evidence which suggests that a person will not learn much from a string of words simply by articulating them. If a person is to learn the information conveyed in the words it appears necessary for him to give the words some form of meaningful representation, and this he does not always do spontaneously. Several experiments have demonstrated that procedures which cause subjects to comprehend the meaning of sentences strongly facilitate learning. For instance, Bobrow and Bower (1969) instructed subjects to compose sentences which were sensible continuations of each sentence they saw. For example, given the sentence "The farmer discovered a diamond," a sensible continuing sentence is "He sold it to a jeweler and used the money to buy a tractor." The subjects in the continuation condition recalled twice as many predicate nouns, given the subject nouns as retrieval cues, than control subjects required to read each sentence aloud three times.

It seems probable that the questions to be answered in a self-instructional program can, under some conditions, get the student to process with understanding the material that precedes the questions (cf. Rothkopf, 1969). If a thematic prompt helps a student fill a blank, he will be more likely to use the prompt, and having used the prompt he may thereby learn more readily. Illustrated below is the form of the thematically prompted sentences investigated in Experiment II.

Before turning red, traffic SIG
are ______.

Under this modified procedure, it was expected that subjects would try to bring to mind a meaningful representation of the rest of the sentence in order to generate the response term. Simply translating the words into implicit speech would not do because, obviously, the mere sound of the words could not evoke the response. It also seemed obvious that response determination would be a potent variable under the new procedure, since it would seem that a prompt would be irrelevant unless it helped evoke the response.

Method

Subjects. The subjects were 56 college students participating to fulfill a class requirement. Approximately two-thirds of the subjects participated in the present study after completing a computer-assisted instruction experiment concerned with feedback procedures. The computer-assisted instruction experiment required approximately 80 minutes to complete. The remaining one-third of the subjects were in a posttest-only group (i.e., no instruction) from the above mentioned study and participated in the present experiment after spending approximately 25 minutes completing the posttest. There was no difference in performance in the present study as a function of amount of time spent in the computer-assisted instruction study. The experiment was run in small groups with the size of the groups ranging from one to six. The subjects were assigned to the experimental conditions by having one member of a pair of subjects call a coin flip. The two groups of subjects were then run simultaneously in different rooms.

Materials and procedure. The materials were 28 of the CVC-word pairs used in Experiment I, with seven pairs representing each combination of levels of associability and response determination. One group saw the pairs embedded in thematically prompted sentences. The second group saw the pairs included within unprompted sentences. In each case, the sentence, without the response term, was projected on a screen for 4 seconds using an overhead projector. At the end of 4 seconds, the plastic overlay was moved slightly to expose the response term (along with the rest of the sentence) for an additional 2 seconds. No overt response was required during item presentation. There was a 3 minute recall interval following the presentation of all 28 items, during which the subject wrote his responses on a sheet listing the CVCs. There were four study trials and four recall intervals, each of which entailed a different random order of items.

Results

Figure 1 shows the mean percentage of correct responses under the two prompting conditions on each of the four trials. The prompted condition was consistently superior to the unprompted condition, $F = 21.19$, $df = 1/54$, $p < .01$. There was also a significant Prompting Condition × Trials interaction, $F = 7.87$, $df = 3/162$, $p < .01$, due to the larger differences between conditions on the first three trials than on the last trial.

The expected Prompting Conditions × Associability interaction appeared in this experiment, $F = 6.50$, $df = 1/54$, $p < .05$. The thematically prompted sentences were relatively more effective with hard-to-associate pairs than unprompted sentences. The Prompting Condition × Associability × Response Determination interaction was significant, $F = 7.69$, $df = 1/54$, $p < .01$. The prompted sentences showed the greatest advantage on hard, determined items and the least advantage on easy, determined items.

As is apparent from Figure 1, there were significant performance gains over trials, $F = 449.42$, $df = 3/162$, $p < .01$. Also significant were the effects of Associability, $F = 35.19$, $df = 1/54$, $p < .01$, and the Associability × Trials interaction, $F = 2.70$, $df = 3/162$, $p < .05$. There was a greater difference between easy and hard items on the first and second trials than on the third and fourth trials. As in Experiment I, the difference between easy- and hard-to-associate pairs was greater when response determination was low than when it was high, $F = 4.78$, $df = 1/54$, $p < .05$.

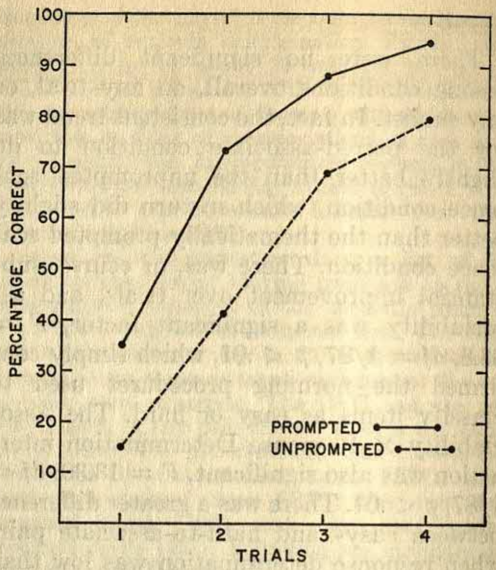


FIG. 1. Mean percentage correct per trial in Experiment II.

The only other significant effect was the Prompting Condition × Trials × Response Determination interaction, $F = 3.09$, $df = 3/162$, $p < .05$, which did not appear to arise from orderly trends in the data and which will, therefore, be left unexplained.

EXPERIMENT III

Experiment II demonstrated that when sentences are first exposed with blanks in place of the response terms people learn CVC-word pairs more readily from thematically prompted sentences than from unprompted sentences. However, the advantage of the thematically prompted sentences may have appeared for a trivial reason. For at least half of the items, those in which the response was determined, the subject could generate the response during the

4-second interval before the response term was displayed. Thus, the cue and the response term were available together for rehearsal and mental elaboration for an average of more than 2 seconds under the prompted condition. Under the unprompted condition, on the other hand, the cue and response were never available together for more than 2 seconds until an item was learned. There is evidence that even when total study time is equated there can be differences among procedures in the amount of effective practice time which influence rate of learning (Cunningham & Anderson, 1968). The purpose of Experiment III was to determine whether the observed superiority of thematically prompted sentences could be attributed to differences in effective practice time.

Method

The subjects were 64 college students participating to fulfill a class requirement. As in Experiment II, approximately one-third of the subjects participating in the study were in the posttest-only group of the previously mentioned computer-assisted instruction experiment. The remaining two-thirds of the subjects participated in the present experiment after spending about 80 minutes on the computer-assisted instruction experiment. Again, there was no appreciable difference in performance on the present study between these two groups of subjects. The materials and procedure for Experiment III were identical to Experiment II except for a change in the unprompted sentence condition. Whereas in the previous experiment the sentences were presented in the form, for instance, "SIG are ______," they were now presented with the blank filled in, for instance, "SIG are *yellow*." That is, rather than seeing the item stem for 4 seconds and then the item stem plus the response for 2 additional seconds, the unprompted group saw the entire sentence for 6 seconds. The prompted group, on the other hand, saw the sentence minus the response term for 4 seconds, and then the sentence including the response term for an additional 2 seconds.

Results

Figure 2 shows the mean percentage of response terms recalled on each of the four trials. Although the difference between the two groups has been attenuated considerably when compared with Experiment II, the prompted group still performed significantly better than the unprompted group, $F = 4.96$, $df = 1/62$, $p < .05$. In addition, prompting condition was involved in one significant interaction, Prompting Condition × Associability × Response Determination, $F = 4.70$, $df = 1/62$, $p < .05$. The prompted group was superior to the unprompted group for every type of item except the easy, undetermined items.

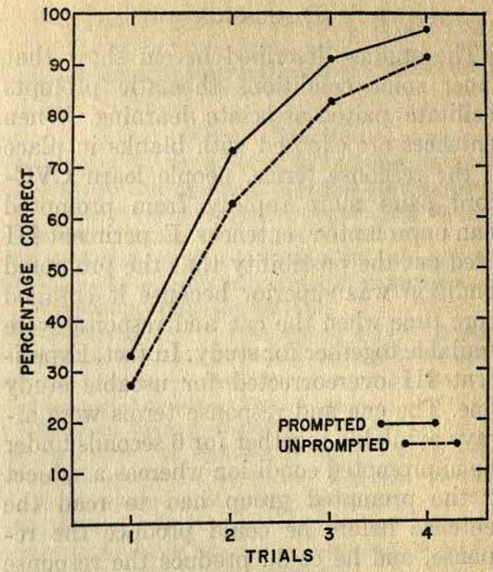


FIG. 2. Mean percentage correct per trial in Experiment III.

As in the first two experiments, trials, $F = 689.46$, $df = 3/186$, $p < .01$, and associability, $F = 39.67$, $df = 1/62$, $p < .01$, resulted in significant effects. Furthermore, trials interacted with associability, $F = 10.27$, $df = 3/186$, $p < .01$, and response determination, $F = 3.96$, $df = 3/186$, $p < .01$. In each case, the differences were greater on the first and second trials than on the third and fourth, undoubtedly due at least in part to a performance ceiling. Nineteen subjects performed perfectly on Trial 3 and 38 did so on Trial 4.

Once again there was a greater difference between easy- and hard-to-associate pairs when response determination was low than when it was high, $F = 10.18$, $df = 1/62$, $p < .01$. Even though it has appeared three times, this interaction is not especially noteworthy since it could well be due to properties of the specific items included within sublists. No other significant effects were obtained.

DISCUSSION

The studies described herein show that under some conditions thematic prompts facilitate paired-associate learning. When sentences are exposed with blanks in place of the response terms, people learn CVC-word pairs more rapidly from prompted than unprompted sentences. Experiment III ruled out the possibility that the prompted condition was superior because it entailed more time when the cue and response were available together for study. In fact, Experiment III overcorrected for useable study time. The cue and response terms were always available together for 6 seconds under the unprompted condition whereas a subject in the prompted group had to read the sentence before he could produce the response, and he could produce the response before the response term appeared in only about half the items anyway. Thus, the prompted condition in Experiment III outperformed the unprompted condition despite having less useable practice time.

Prompted sentences were expected to produce greater facilitation with hard-to-associate pairs than with easy-to-associate pairs. While the effect was small, the predicted interaction between prompting condition and associability was significant ($p <$ .05) in Experiment II, and in the right direction but not significant ($p < .18$) in Experiment III.

The sentence-with-blank procedure was employed in Experiments II and III because it was believed that this procedure would induce meaningful processing of the sentences. The fact that thematic prompts proved facilitative in these experiments, but not in Experiment I where entire sentences were exposed, is consistent with this hypothesis. There are other explanations, however. In Experiment I the sentences were presented at a 4-second rate, which may not have allowed subjects enough time to completely process the thematic prompts. Or, it is possible that the different outcomes of the three experiments were somehow due to the fact that subjects were run individually in Experiment I but in groups in the other two studies.

There are at least two explanations for the facilitative effects of thematic prompts observed in the present studies. The first is that the prompt furnished a mediator, a link between the cue and the response. The second is that the prompt provided a stable, meaningful encoding of the cue (cf. Martin, 1968). Earlier it was argued that response determination represents the capacity of a prompt to mediate a cue and a response. If this is true, the fact that response determination was an impotent variable implies that the prompts were not functioning as mediators. The subjects may have used the prompts to encode the CVCs, and then dreamed up their own mediators to link the encoded CVCs with the response terms.

Auxiliary data from Experiment I bear on the cue encoding hypothesis. Following the criterion trial, the CVCs were presented again. The subjects were asked to provide the English word that each CVC stood for. For instance, they were expected to say "signal" for SIG and "shoes" for SUS. The prompted sentence group gave the encoding planned by the authors 70% of the time, whereas the figure was 12% for the unprompted sentence group and the paired-associate group. These data are difficult to interpret. On the one hand, the subjects in the prompted sentence group did learn most of the cue encodings which the authors intended. Yet the prompted group learned the pairs no more quickly than the other groups which would, therefore, suggest that the facilitation that accompanied thematic prompts cannot be explained in terms of cue encoding. On the other hand, the norming sample gave the intended encodings for better than 95% of the CVCs, indicating that the average percentage for the prompted sentence group could have been considerably higher. Furthermore, there was a correlation of .26 between total correct responses and the number of cue encodings which matched the authors' encodings. This may indicate that subjects who used the encodings contained in the prompts learned the pairs more readily. In sum, while not clearly supported on the basis of the present data, the cue encoding hypothesis remains viable.

Consider now the mediation hypothesis. The cue encoding hypothesis is attractive to explain data from studies employing CVCs as cues, but it probably cannot explain the results of the experiments in which the cues have been common nouns, since presumably the subject can use a common noun without transformation or elaboration. Some form of mediational explanation seems necessary to account for the facilitative effects of providing verbs or prepositions to link pairs of nouns, or instructing subjects to generate their own sentences to link them. It is parsimonious to believe that thematic prompts also played a mediational role in the present experiments.

As already mentioned, the fact that response determination made little difference is difficult to square with the mediation hypothesis. It would certainly seem that, given the cue and a good mediator, the subject should be able to produce the response. And, it would seem that a mediator is poor if, given it and the cue, the response is indeterminate. Nonetheless, whether or not an experimenter-provided mediator, or thematic prompt, determined the response proved unimportant in the present experiments. This fact can be accommodated within a reformulation of the mediation hypothesis which the authors have adapted from Bobrow's and Bower's (1969) theorizing about sentence comprehension. The idea is that what is important is the *process* of conceiving a link between the cue and the response rather than the *product* that results from such activity. According to this view one mediating link is as good as another, provided only that the subject conceives the link while learning. By "conceives a link," it is meant that he brings to mind a relationship between the thing or event named (suggested) by the stimulus term and the thing or event named (suggested) by the response term. While obviously speculative, this hypothesis is consistent with the major facts of the present studies and with the principal findings of other recent research (Bobrow & Bower, 1969). The general implication is that the search for variables which influence the kind of processing people give to verbal materials is likely to be more fruitful than the search for variables that distinguish effective from ineffective mediators.

REFERENCES

ANDERSON, R. C. Control of student mediating processes during verbal learning and instruction. *Review of Educational Research,* 1970, **40,** 349–369.

BOBROW, S. A., & BOWER, G. H. Comprehension and the recall of sentences. *Journal of Experimental Psychology,* 1969, **80,** 455–461.

CUNNINGHAM, D. J., & ANDERSON, R. C. Effects of practice time within prompting and confirmation presentation procedures on paired associate learning. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7,** 613–616.

KIESS, H. O., & MONTAGUE, W. E. Natural language mediators in paired associate learning. *Psychonomic Science,* 1965, **3,** 549–550.

KULHAVY, R. W. Natural language mediators and paired associate learning in college students. *Psychological Reports,* 1970, **26,** 658.

MARKLE, S. M. *Good frames and bad.* (2nd ed.) New York: Wiley, 1969.

MARTIN, E. Stimulus meaningfulness and paired-association transfer: An encoding variability hypothesis. *Psychological Review,* 1968, **75,** 421–441.

MONTAGUE, W. E., ADAMS, J. A., & KIESS, H. O. Forgetting and natural language mediation. *Journal of Experimental Psychology,* 1966, **72,** 829–833.

MONTAGUE, W. E., & KIESS, H. O. The associability of CVC pairs. *Journal of Experimental Psychology Monograph,* 1968, **78**(2, Pt. 2).

MONTAGUE, W. E., & WEARING, A. J. The complexity of natural language mediators and its relation to paired-associate learning. *Psychonomic Science,* 1967, **7,** 135–136.

ROHWER, W. D., JR. *Social class differences in the role of linguistic structures in paired-associate learning.* (Final report, USOE Cooperative Research Program Project No. 5-0605.) Berkeley: University of California, 1967.

ROTHKOPF, E. Z. Some conjectures about inspection behavior in learning from written sentences and the response mode problem in programed instruction. In R. C. Anderson, G. W. Faust, M. C. Roderick, D. J. Cunningham, & T. André (Eds.), *Current research on instruction.* Englewood Cliffs, N. J.: Prentice-Hall, 1969.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia University Press, 1944.

UNDERWOOD, B. J. Stimulus selection in verbal learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning.* New York: McGraw-Hill, 1963.

(Received April 7, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 4, 322–327

RETENTION AND TRANSFER OF MEDIATION SET IN PAIRED-ASSOCIATE LEARNING OF EDUCABLE RETARDED CHILDREN†

DOROTHEA ROSS[1]

Stanford University School of Medicine

Measures of ability to learn paired associates and to use mediational links were obtained on young educable mentally retarded subjects in a laboratory setting. The experimental group subjects ($n = 19$) were given long-term mediation training within the context of a special music program which emphasized the principle and utility of mediational links. The control group subjects ($n = 19$) participated in a traditional primary music program. Postexperimental measures showed retention and transfer of the mediation training by the experimental group with marked improvement in their nonverbal ($p < .001$) and verbal ($p < .0005$) paired-associate performance.

The mentally retarded child is deficient in the use of verbal mediators. The fact that this deficiency does not disappear with age as it does in children of normal intelligence (Luria, 1960; Reese, 1962) has led to investigations of the efficacy of mediation training. Although the retardate can be taught to use mediating processes, the short-term training procedures used have not resulted in the retention of either a specific mediational strategy or a mediation set (Jensen, 1971; Milgram, 1967, 1968).

While any procedure that facilitates learning in retardates is of interest, the critical question concerns the retention and transfer of verbal habits (Keppel, 1968). The failure to retain a mediation set may be due to an inability to understand the function of the mediational strategies that are supplied (Milgram, 1968). Insufficient practice may also be a factor, in view of the extensive repetition needed by the retardate in acquiring new verbal responses.

However, even if retention of mediation set were to occur, the specificity of the mediational strategies and the homogeneity of the training materials used by previous investigators impose limits on the utility of the learning. A training procedure having greater potential should emphasize the *principle* of mediational links in order that the retardate may effectively use available mediational opportunities.

In the present study, educable mentally retarded children participated in a mediation training program that differed markedly from those of previous investigators (Jensen, 1971; Jensen & Rohwer, 1963; Milgram, 1968). The salient characteristics of the program were long-term group training, tasks that demonstrated the utility of mediational links, and emphasis on understanding the principle of mediation.

The purpose of the study was to test the efficacy of this mediation training as well as to demonstrate the generality of the response patterns learned.

METHOD

General Procedure

Preexperimental measures of ability to learn paired associates and of the use of verbal mediators were obtained from educable mentally retarded children. The experimental group was then given mediation training within the context of a 6-week special music program and the control group spent equal time in a traditional primary music program. Postexperimental measures of ability to learn paired associates and of the use of verbal mediators were obtained.

Subjects

The 38 subjects ($M = 23$, $F = 15$) ranged in chronological age from 6 years 2 months to 10 years 6 months; in IQ from 51 to 80 on the Stanford-Binet Intelligence Test, Form L-M; and in mental age from 3 years 9 months to 8 years 4 months. The subjects were all of middle-class background and were enrolled in four primary classes for educable mentally retarded children. The four classes did not differ in chronological age, IQ, mental age, sex, years in school, preexperimental ability to learn paired associates, and verbal mediation scores. In order that the music program could be conducted as a regular classroom activity, the four classes were randomly assigned by class to experimental and control groups. The two experimental ($n = 19$) and two control classes ($n = 19$) did not differ in chronological age, IQ, mental age, sex, years in school, preexperimental ability to learn paired associates, and verbal mediation scores.

---

[1] The author wishes to express her appreciation to Geoffrey Keppel for his critical reading of this paper.

Requests for reprints should be sent to Dorothea Ross, Department of Pediatrics, Stanford University School of Medicine, Stanford, California 94305.

Research Assistants

The research assistants were undergraduates in psychology and education who were trained to teach the music programs and were randomly assigned one to each of the four classes. Neither the research assistants nor the classroom teachers knew about the paired-associate learning task. The classroom teachers did not participate in any phase of the music programs.

Two other research assistants collected the paired-associate learning measures. To avoid transfer occurring as a result of similarity of task setting or association of personnel, these research assistants tested individual subjects in experimental rooms, were not in the schools during the music program, and had no apparent connection with the music activities or personnel.

Paired-Associate Materials

Preexperimental and postexperimental measures of ability to learn paired associates were obtained with a motor and a verbal task, each consisting of two equivalent lists of seven paired associates (Motor Task—Sets 1 and 2, Verbal Task—Sets A and B).

The purpose of the Motor Task was to familiarize the subjects with the experimental task (Prehm, 1968). The stimulus items were arithmetic symbols, for example, a division sign, mounted on 6 × 9 inch cards and the response items were motor responses, for example, one hand closed. All subjects learned the Motor Task without difficulty.

For the Verbal Task, the stimulus items were abstract symbols, for example, an octagon, and the response items were colored pictures of familiar objects, for example, a horse. Both sets of items were mounted on 8 × 11 inch cards. Prior to the experiment, each subject was presented with the 14 response pictures in Sets A and B and with foils, to ensure that he could name the pictures.

Procedure

For the preexperimental paired-associate learning measures the experimenter brought each subject to an experimental room and said:

> We are going to play a game called Signals. Do you know what a signal is? (Pause) A signal tells you what to do. When your father is driving a car and he comes to a red traffic light, he stops, doesn't he? The red light is a *signal* that tells him to stop. What signal does a policeman make that tells your father to stop? (Pause) That's right, he holds up his hand and that is a signal to stop. Your teacher uses this signal when she wants you to be very quiet. She puts her finger to her mouth like this [the experimenter demonstrates]. That signal tells you to be quiet.
>
> Now we are ready to play. I think you will be good at this game. You already know some signals. I will put a signal picture here [the experimenter points to the stimulus stand] and when you see it you must answer with your hands. First I will show you each of the signal pictures and the answers that you are to make with your hands, then you can have a turn.

The experimenter demonstrated the paired-associate learning procedure with four practice cards and then gave each subject the Motor Task.

Motor Learning

The standard anticipation procedure was used with a 4-second anticipation interval followed by a 4-second simultaneous presentation of the stimulus and the response items and then a 4-second interpair interval. Prior to the first anticipation trial (Trial 1) the subject observed the pairs without being required to respond (exposure trial). The order of presentation of the two sets was counterbalanced. Four orders of presentation of each set were used to minimize serial learning. A pair was considered to be learned after four correct, though not necessarily consecutive, anticipations. The pair was then removed and the remaining pairs were presented according to the prearranged schedule until each item had been learned. The trials to criterion will be referred to as the motor original learning.

Motor Retention

From 7 to 9 days after the motor original learning measure, retention was tested by recall and by relearning. The motor recall test consisted of three complete trials in which only the stimulus items were presented and the subject's score was the total number of correct anticipations. The procedure for the motor relearning test was similar to that of the original learning test except that the instructions were brief and there was no exposure trial.

Verbal Learning

One week after the Motor Task retention measures the same general procedure was used for the Verbal Task. The learning measures were correct anticipations on Trial 1 (Verbal Trial 1) and trials to criterion of four correct anticipations of each pair (verbal original learning). No retention tests were conducted.

Observed Mediational Activity

During the Verbal Task trials, trained observers recorded spontaneous reports of the use of mediators, other audible comments, and instances of whispering and lip movements.

Mediational Inquiry

Immediately following the Verbal Task, the experimenter showed the subject each of the seven pairs in order of correct anticipation and said, "Look, here's the first (second, third, etc.) signal you got right. That was good work. How did you get this signal right so quickly?" All responses were recorded but ambiguous ones were also queried. The total number of mediational links reported was the subject's verbal mediation score.

Classification of Responses

Two judges did independent blind ratings of the 38 subjects' responses to the inquiry. An agreement was counted whenever both judges assigned a response to the same general class (mediated or nonmediated) and, in the case of mediated responses, to the same subclass (chained associative links, rhyming, sentence, conjunction, preposition). Rater reliability determined by the percentage agreement method was 98%.

Music Programs

One week after the completion of the preexperimental paired-associate learning measures, the experimental group participated for 6 weeks in a music program which provided mediation training. The control group spent equal time in a traditional primary music program.

The experimental group program was adapted from the Kodaly-Richards Music Program (Richards, 1964). In this program, the child first learns to read rhythm patterns and simple melodies from meaningful pictures and then gradually makes the transition to standard musical symbols on a treble clef. The program is a series of associative learning tasks: The child must learn to associate series of stimulus items with the appropriate motor and verbal responses. For example, he learns to clap loudly when he sees a picture of an elephant walking and softly for a mouse walking; he learns to verbalize quickly to a series of quarter notes and slowly to a single whole note. Mediational links to facilitate the acquisition of these tasks were supplied by the research assistants at the beginning of the training period and later were formulated by the subjects. The research assistants taught a variety of mediational strategies, emphasized the utility of mediators, and praised the subjects for using supplied and formulated mediators all within the context of the music training.

The control group program was the Primary Music Program of Songs (Watters, 1965), a traditional primary program of simple rhythm and song activities. This program is a rote learning task rather than one that depends on the understanding or use of links to remember sequences.

Both music programs were taught by the research assistants in the subjects' own classrooms for three 30-minute sessions a week for 6 weeks.

Postexperimental Measures

From 7 to 9 days after the music training, the preexperimental measures of nonverbal and verbal paired-associate learning and of use of verbal mediators were again obtained on all 38 subjects.

RESULTS

On the preexperimental measures, the two groups did not differ. On the posttesting, however, the experimental group performed better than the control group on all six dependent variables. These results are shown in Table 1. Within-group comparisons showed that the experimental group improved on the motor original learning ($t = 3.64, p < .002$, two-tailed) and verbal original learning ($t = 6.15$, $p < .0005$, two-tailed) whereas the control group did not improve on either of these measures.

The subjects reported the following types of links in the preexperimental ($n = 6$) and postexperimental ($n = 52$) inquiry sessions: conjunctions ($n = 13$), prepositions ($n = 8$), sentences ($n = 20$), rhyming ($n = 4$), and chained associative links ($n = 13$). In the postexperimental sessions, 16 subjects in the experimental group used mediational links compared to 3 subjects in the control group. There was no evidence from either specific pairs or total scores that qualitative aspects of the subjects' reported mediators were related to efficiency of learning.

DISCUSSION

It is clear that the educable mentally retarded child can learn strategies to relate items to one another and that he can initiate and use these strategies with a consequent improvement in his performance on associative learning tasks.

The subjects who had learned to use mediational strategies in a classroom activity subsequently used them effectively in an unrelated laboratory task, a convincing

TABLE 1

POSTEXPERIMENTAL COMPARISONS OF THE EXPERIMENTAL AND CONTROL GROUPS ON MOTOR AND VERBAL MEASURES

Group	*n*	*M*	*SD*	*t*	*M*	*SD*	*t*
		Motor original learning			Verbal Trial 1		
Experimental	19	5.53	1.39	3.69***	3.79	.92	1.97*
Control	19	8.05	2.55		3.10	1.15	
		Motor recall			Verbal original learning		
Experimental	19	14.26	2.68	2.66**	7.00	1.92	4.49****
Control	19	11.21	4.06		10.21	2.35	
		Motor relearning			Verbal mediation		
Experimental	19	5.21	.63	3.80***	2.47	2.06	4.70****
Control	19	6.84	1.71		.26	.44	

* $p < .05$, two-tailed test.
** $p < .01$, two-taled test.
*** $p < .001$, two-tailed test.
**** $p < .0005$, two-tailed test.

demonstration of the retention and transfer of mediation set. Additional evidence of the generality of this response pattern was provided by teachers' reports that the subjects' use of mediational strategies transferred to other activities. Of interest here are the procedures in the training program that activated the mediating process (Jenkins, 1963).

The music program (Richards, 1964) has the inherent capacity to elicit associative arousal and consequent mediational activity. The rhythm tasks vary from high-strength to low-strength associates, thus providing a graded series of opportunities for seeing the utility of mediators and for developing an understanding of the principle of mediation. The use of group training allowed the subjects to see other subjects (models) frequently rewarded for formulating mediators. That this modeling procedure increased the frequency with which the subjects subsequently exhibited mediational responses is consistent with previous studies of modeling procedures with educable mentally retarded children (Ross, 1968, 1970a).

The results of a current study (Ross, 1970b) support the conclusion that the critical factors for retention and transfer are the salient characteristics of the training program (e.g., emphasis on utility) rather than the specific task (e.g., music program).

Previous investigators have found that better organized and more meaningful links such as sentences result in more facilitation of ability to learn paired associates than do prepositions, conjunctions, or labeling (Rohwer, 1966). In the present study, there was no evidence from either specific pairs or total scores that qualitative aspects of the subjects' reported mediators were related to efficiency of learning. Effective mediational strategies ranged from simple conjunction links to more complex sentence structures. An irregular series of jagged lines (stimulus) and an Easter egg (response) elicited the following mediators, all of which were associated with consecutive correct responses on Trials 1 through 4:

> Well, this [stimulus] is an up-down and I say *up-down and Easter egg* and every time I see the up-down I say *Easter egg.*
>
> This here hopper thing [stimulus] is a thing going hop hop to the big egg so I say *hop to the egg, hop to the egg* a lot of times and every time the hop picture comes then the egg picture comes the very next.
>
> Well, see that is the way the Easter bunny goes hopping, so I say *the Easter bunny brings me eggs*

and when the bunny picture comes I say *Easter egg* real fast.

Now this [stimulus] looks like the way you jump along, see, up and down and up and down. Now, who goes jumping like that? Not Santa Claus, no, the Easter bunny. So when this [stimulus] comes I say *jump, jump* and that makes me say *Easter bunny* and then I know its egg.

One explanation for the finding that the better organized links were not related to more efficient learning is that in some cases the reported formula was identical to the first overt mediating response while in other cases the reported formula was the last of several modifications of the original mediating response. The implication of this finding is that the procedure of obtaining reports of mediators at the conclusion of a paired-associate learning task may yield information about processes operating at the time of retrieval rather than about the associative bond formed at the time of storage. Rohwer, Shuell, and Levin (1967) have argued that it is the bond formed at the time of storage that is the critical factor in mediational facilitation.

The data provide some support for a production deficiency (Maccoby, 1964). The preexperimental inquiry elicited only six reported mediators and direct observation revealed few instances of detectable verbalization other than simple naming. Failure to report mediators cannot be taken as sufficient evidence that the subject has not used them but the absence of detectable verbalization does suggest a lack of mediational activity. The educable mentally retarded child is usually unable to engage in completely covert verbal coding and rehearsal in new learning situations (Denny, 1964). Further support for a production deficiency is provided by the fact that throughout the training program the subjects effectively used supplied and spontaneously formulated links.

This study is the first of a series of investigations of the efficacy of intensive training for the educable mentally retarded child in the set of subskills that are prerequisite to effective learning and retention. The assumption is that the performance deficit of the educable mentally retarded child is due to his inability to use his intellect effectively and that with training he can become competent in cognitive skills with a consequent improvement in performance. The present study supports this assumption and further suggests that the educable mentally retarded child's inability to learn certain cognitive tasks is not an irreversible result of his slower rate of intellectual development but is due instead to the use of inappropriate training methods.

REFERENCES

DENNY, M. R. Learning and performance. In H. Stevens & R. Heber (Eds.), *Mental retardation.* Chicago: University of Chicago Press, 1964.

JENKINS, J. J. Mediated associations: Paradigms and situations. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning.* New York: McGraw-Hill, 1963.

JENSEN, A. R. The role of verbal mediation in mental development. *Journal of Genetic Psychology,* 1971, **118,** 39–70.

JENSEN, A. R., & ROHWER, W. D., JR. The effect of verbal mediation on the learning and retention of paired-associates by retarded adults. *American Journal of Mental Deficiency,* 1963, **68,** 80–84.

KEPPEL, G. Comments on Professor Prehm's paper. In H. J. Prehm, L. A. Hamerlynck, & J. E. Crosson (Eds.), *Behavioral research in mental retardation.* Eugene: University of Oregon Press, 1968.

LURIA, A. R. *The role of speech in the regulation of normal and abnormal behavior.* Washington, D.C.: United States Government Printing Office, 1960.

MACCOBY, E. E. Developmental psychology. *Annual Review of Psychology,* 1964, **15,** 203–250.

MILGRAM, N. A. Retention of mediation set in paired-associate learning of normal children and retardates. *Journal of Experimental Child Psychology,* 1967, **5,** 341–349.

MILGRAM, N. A. The effect of verbal mediation in paired-associate learning in trainable retardates. *American Journal of Mental Deficiency,* 1968, **72,** 518–524.

PREHM, H. J. Rote verbal learning and memory in the retarded. In H. J. Prehm, L. A. Hamerlynck, & J. E. Crosson (Eds.), *Behavioral research in mental retardation.* Eugene: University of Oregon Press, 1968.

REESE, H. W. Verbal mediation as a function of age level. *Psychological Bulletin,* 1962, **59,** 502–509.

RICHARDS, M. H. *Threshold to music.* Palo Alto: Fearon, 1964.

ROHWER, W. D., JR. Constraint, syntax and meaning in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5,** 541–547.

ROHWER, W. D., JR., SHUELL, T. J., & LEVIN, J. R.

Context effects in the initial storage and retrieval of noun pairs. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 796–801.

Ross, D. M. *The use of games to facilitate the learning of basic number concepts in preschool educable mentally retarded children.* (United States Office of Education, Project No. 6-2263) Stanford, Cal.: Stanford University, School of Medicine, 1968.

Ross, D. M. The effect of psychological attachment to a film-mediated model on learning in retarded children. *American Journal of Mental Deficiency*, 1970, **74**, 701–707. (a)

Ross, D. M. The storage and use of previously developed mediators in retarded children. Unpublished manuscript, Stanford University, 1970. (b)

WATTERS, L. E. *The magic of music.* Boston: Ginn, 1965.

(Received April 15, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 4, 328-332

ADJUSTMENT OF ANGLO-AMERICAN AND MEXICAN-AMERICAN PUPILS IN SELF-CONTAINED AND TEAM-TEACHING CLASSROOMS[1]†

LINDA SCHMIDT AND JUNE GALLESSICH[2]
University of Texas

Pupil adjustment in self-contained and team-teaching classrooms was compared. The subjects were first-grade ($n = 160$) and sixth-grade ($n = 382$) pupils from experimental team-teaching schools (one predominately Anglo-American and one predominately Mexican-American) and self-contained classroom schools with comparable populations. Anxiety was measured with the Picture Anxiety Test (first-grade subjects) and the Phillips Anxiety Test (sixth-grade subjects); pupil evaluation of teachers was measured with the Pupil Observation Survey Report. Analysis of variance was used to study relationships between school adjustment and three variables: teaching organization, ethnic origin, and sex. A number of significant relationships were found.

An especially controversial issue in team teaching is the impact on the mental health of young children of the loss of the single "mother hen" teacher model (Anderson, 1966; Perkins, 1966). And yet, few studies of this aspect of team teaching have been reported. Carbone (1963) and Lambert, Goodwin, and Wiersma (1965) found no significant relationship between classroom organization and emotional adjustment. Phillips (1966), however, discovered a significant reduction in anxiety in nongraded, team-taught third, fourth, and fifth graders.

Another team-teaching issue has been raised by educators who work with Mexican-American pupils. These educators question the effect of the team approach upon Mexican-American children, who are separated from the mainstream culture by economic and language disparities. The transition from home to school is a more drastic change for Mexican-Americans than for Anglo-Americans and is, therefore, a time of greater insecurity (Manuel, 1965). The Mexican-American seldom knows English when he enters school; his culture shock is accentuated by his typically Anglo teacher. Manuel (1965) found that the initial adjustment to school is a critical factor in the later school achievement of these children.

The purpose of the present study was to investigate the relationship between school adjustment and three variables: teaching organization, ethnicity, and sex.

METHOD

Subjects

The subjects were pupils from four elementary schools in a central Texas city. The two experimental schools were organized for team teaching; the population of one of these schools was predominately Anglo-American and the other was predominately Mexican-American. The populations of the two self-contained control schools were comparable to the experimental schools in ethnicity and socioeconomic status. First- and sixth-grade pupils were selected for the study. The first-grade sample of 160 subjects was composed of 40 children (20 males and 20 females) randomly selected from each of the four schools. The entire sixth-grade populations of the four schools, a total of 382, were the subjects of the older sample: 133 subjects from the Anglo-American team-teaching school, 92 subjects from the Mexican-American team-teaching school, 94 subjects from the Anglo-American self-contained school, and 63 subjects from the Mexican-American self-contained school.

Instruments

Instruments designed to measure anxiety and pupil perception of teacher were used.

Anxiety of first graders was measured with the

---

[1] This work was supported by the Research and Development Center for Teacher Education, the University of Texas at Austin, under United States Office of Education Contract 6-10-108.

[2] The authors wish to thank the Austin Independent School District and principals K. I. Isbell and R. D. Evans for making this investigation possible.

Requests for reprints should be sent to June Gallessich, Counseling—Psychological Services Center, University of Texas, Austin, Texas 78712.

Picture Anxiety Test, a modification of an earlier form developed by Temple and Amen (1944) and Dorkey and Amen (1947). The revised Picture Anxiety Test is a set of 22 pictures of children in potentially anxious situations at home and school.[3] For example, in one picture a child is standing at the teacher's desk. The child in each picture has no mouth. Below each picture are two faces of children, one happy and one sad. The subjects are asked to look at the picture situation and below it mark the face that the child in the picture would have. The original version of this test was found to have a Pearson r of .67 on an 8-week test–retest with 95 first graders. The odd-even internal consistency index (Spearman-Brown correction) with nursery children was .95. The test was found to be positively correlated ($r = .57$) with teacher ratings of anxiety in a small nursery population (Dorkey & Amen, 1947; Temple & Amen, 1944).

Anxiety of sixth graders was measured by the Phillips Anxiety Test (Phillips, 1966) which consists of 33 items such as:

> Are you often worried that the teacher will scold or punish you?
>
> Do you worry a lot *while* you are taking a test?
>
> Do you often have the fear that other children might think you dumb?

The Pupil Observation Survey Report (Veldman & Peck, 1963, 1967) was used to measure pupil perception of teacher. The primary form used for the first-grade subjects consists of 35 questions such as:

> Does she teach things that are fun?
>
> Have you learned a lot from her?

The 38-item intermediate form of the Pupil Observation Survey Report used with the sixth-grade subjects contains such items as:

> She usually looks on the bright side of things.
>
> I wish all my teachers were like her.
>
> She has made her subject alive and interesting for me.

A measure of the pupil's general evaluation of the teacher is derived from the Pupil Observation Survey Report. A reliability index of .69 was found in an 8-week test–retest with 80 first graders.

Procedure

The first-grade subjects were tested in groups of five to eight (mixed sex) by a bilingual examiner who orally administered the tests to all first-grade subjects. Spanish-speaking subjects were given the tests in both English and Spanish. One author was present at each testing session to assist in administration. The entire sixth-grade population of each of the four schools was tested in a single administration in the school cafeteria. The same persons administered the tests to the sixth-grade subjects of all four schools. A proctor was provided for every 15 subjects.

Analysis of variance technique was used to investigate relationships between school adjustment, teaching organization, ethnicity, and sex.

RESULTS

Anxiety Analyses (Picture Anxiety Test and Phillips Anxiety Test)

Cultural effect. Mexican-American first-grade and sixth-grade subjects reported significantly higher anxiety levels than Anglo-American subjects. The mean of the Anglo-American first graders on the Picture Anxiety Test was 13.93, while the mean of the Mexican-American first graders was 16.67 ($p < .001$). The Mexican-American sixth-grade mean of 19.91 on the Phillips Anxiety Test was significantly higher ($p < .002$) than the Anglo-American mean of 16.08.

Organizational effect. First-grade subjects in self-contained classrooms appeared more anxious (Picture Anxiety Test) than first-grade subjects in team-teaching settings ($p < .01$). Sixth-grade subjects in self-contained classrooms were also more anxious ($p < .05$) than team-taught sixth graders, according to their scores on the Phillips Anxiety Test.

Sex effect. The mean anxiety score (Picture Anxiety Test) of the total male first-grade sample did not differ significantly from the female mean. But sixth-grade females were found to be more anxious than males ($p < .03$) when their scores on the Phillips Anxiety Test were compared.

Culture × Organization. The anxiety level of the two ethnic groups did not differ significantly between the two organizations in the first grade. But the anxiety level reported by Mexican-American sixth graders in self-contained classrooms was significantly higher ($p < .01$) than the anxiety reported by Mexican-American team-taught sixth-grade subjects. Sixth-grade Anglo-American subjects in the two teaching organizations reported similar levels of anxiety.

Sex × Organization. First-grade females in self-contained classrooms reported significantly higher anxiety than first-grade females in team teaching ($p < .02$). First-grade males in the two teaching organizations reported similar levels of anxiety. No significant interaction between sex and organization was found in the sixth-grade subjects.

---

[3] The instrument is available from the authors upon request.

Culture × *Sex*. No significant interaction between culture and sex was found in either the first- or the sixth-grade groups.

Analyses of Pupil Evaluation of Teacher (Pupil Observation Survey Report)

Cultural effect. First-grade Mexican-American subjects perceived their teachers less favorably than their Anglo-American counterparts. The Pupil Observation Survey Report mean for Anglo-Americans was 32.44 and for the Mexican-Americans, 29.11 ($p < .0001$). There was no significant difference between Mexican-American and Anglo-American sixth graders' evaluations of teachers.

Organization effect. No organization effect upon teacher evaluation was found in the first grade, but sixth-grade subjects in self-contained classes ($\bar{X} = 28.8$) evaluated their teachers more favorably ($p < .0001$) than students in team-teaching schools ($\bar{X} = 25.99$).

Sex effect. No sex differences in first-grade subjects' perceptions of teachers were found, but sixth-grade males evaluated their teachers much less favorably than did females of that grade ($p < .001$).

Culture × *Organization*. The teacher evaluations by the two ethnic groups did not differ significantly between the two teaching organizations in the first grade; but in the sixth grade a significant Culture × Organization interaction was found. Mexican-American sixth-grade subjects perceived their teachers equally well in the two organizations, but Anglo-American students in team-taught classes reported a much less favorable opinion of their teachers than Anglo-Americans in self-contained classrooms (Pupil Observation Survey Report, $p < .00001$).

Sex × *Organization*. No significant Sex × Organization interactions were found in the first- or sixth-grade levels.

Culture × *Sex*. No significant interactions were found.

DISCUSSION

Cultural Effect on Adjustment

The anxiety level reported by Mexican-American subjects of both grade levels and in both teaching organizations was significantly higher than the anxiety reported by Anglo-American subjects. First-grade Mexican-American subjects viewed their teachers less favorably than did Anglo-American first-grade subjects, but this cultural difference did not appear in sixth-grade subjects. School experience and age appear to have a positive effect on the Mexican-American pupil's impression of his teacher, but these variables do not seem to modify his relatively high anxiety level.

The cultural differences in anxiety may stem from disparate attitudes toward admission of negative feelings. Latins are considered more frank than Anglo-Americans in expressing negative ideas and feelings; Dohrenwend and Dohrenwend (1969) found Puerto Ricans admitted more so-called undesirable symptoms than Irish, Jews, and Negroes. These investigators hypothesized that Puerto Rican judgment of social undesirability is less severe than that of the other groups. The anxiety variance in the present study could be attributed to this "admissions phenomenon"; it could then be inferred that the Mexican-American subjects were not necessarily more anxious but simply more candid than the Anglo-American subjects.

Organizational Effect on Adjustment

The significant organizational effects were found in interaction analysis to be linked to sexual and cultural differences. The heightened anxiety of first-grade subjects in self-contained classrooms was found to be related to sex; first-grade females were much more anxious in self-contained settings than in team teaching, while males' anxiety scores were unaffected by teaching organization. This difference is probably attributable to sexual differences in developmental needs. Boys are beginning at age six to turn away from females while girls are concerned with identification with the feminine adult role (Fenichel, 1945). Possibly the adjustment of first-grade girls, who are ordinarily eager to please "teacher," is enhanced by the availability of several female teacher models. Their chances of finding an approving teacher would appear to be maximized in the team setting. Team teachers may be more pleasant than self-contained classroom teachers. Investigators of team teaching report that teachers who are frequently exposed to colleagues may tend to be more careful than self-contained classroom teachers to present themselves positively to their pupils (*Dundee Team Teaching Project,* 1965; Lortie, 1964).

The comparatively high anxiety of sixth-grade subjects in self-contained classrooms was discovered in interaction analysis to be associated with culture. Mexican-Americans in self-contained classes reported significantly higher anxiety than Mexican-Americans in team teaching; but anxiety scores of Anglo-American sixth-grade subjects were not affected by teaching organization. Yet the teacher perceptions reported by Anglo-American sixth-grade subjects—but not by Mexican-Americans —were significantly affected by teaching organization. These puzzling results are difficult to explain. They are in part understandable when related to observations of the team of teachers who taught the Anglo-American sixth-grade sample. The team was an especially unstable one and had lost 11 of its 12 teachers during the 3 years preceding this investigation. Undoubtedly there was a contagion effect, with the pupils responding to the uneasiness and strife clearly visible in their team. The extremely negative teacher evaluation by this group of subjects was probably a result of conflicts within the teacher team. It is interesting to note that the negative perceptions of the team were not accompanied by heightened anxiety in this group of Anglo-American subjects.

Sex Effect on Adjustment

The school adjustment of females was found in this study to be more variable than male adjustment. The lowered anxiety of first-grade girls in the team setting has been discussed. Female sixth graders—regardless of school organization—reported more anxiety than did the males of this grade. Yet these female subjects perceived their teachers more favorably than did their male peers. Perhaps the greater anxiety of the females is related to the physical, psychological, and social adjustments of puberty which result in concerns typically faced by girls in the sixth grade but not faced by boys until the seventh or eight grade. The females' favorable teacher perceptions are probably related to the identification needs already discussed.

The major objective of this study was to assess the effect of team teaching upon pupil adjustment. The data indicate that team teaching is not detrimental to the elementary grade children of this study—regardless of ethnicity or sex—and may in fact be advantageous for some children. There is one exception to this conclusion, however. The group of Anglo-American pupils taught by a conflicted team of teachers reported an extremely negative perception of their teachers. Probably the traditional self-contained classroom approach which minimizes collaboration as well as opportunities for teacher conflicts is preferable to an obviously disrupted team.

REFERENCES

ANDERSON, R. H. *Teaching in a world of change.* New York: Harcourt, Brace, and World, 1966.

CARBONE, R. F. A comparison of graded and nongraded elementary schools. In J. M. Seidman (Ed.), *Educating for mental health.* New York: Crowell, 1963.

DOHRENWEND, B. P., & DOHRENWEND, B. S. *Social status and psychological disorder.* New York: Wiley, 1969.

DORKEY, M., & AMEN, E. W. A continuation study of anxiety reactions in young children by means of a projective technique. *Genetic Psychological Monograph,* 1947, **35,** 139–183.

Dundee team teaching project. New York: Columbia University, Teachers College, Institute of Field Studies, 1965.

FENICHEL, O. *Psychoanalytic theory of neurosis.* New York: W. W. Norton, 1945.

LAMBERT, P., GOODWIN, W. L., & WIERSMA, W. A. A study of the elementary school teaching team. *Elementary School Journal,* 1965, **66,** 28–34.

LORTIE, D. C. The teacher and team teaching: Suggestions for long-range research. In J. T. Shaplin & H. F. Olds (Eds.), *Team teaching.* New York: Harper & Row, 1964.

MANUEL, H. T. *Spanish-speaking children of the Southwest.* Austin: University of Texas Press, 1965.

PERKINS, B. Team teaching. In F. Guggenheim & C. Guggenheim (Eds.), *New frontiers in education.* New York: Grune & Stratton, 1966.

PHILLIPS, B. *An analysis of causes of anxiety among children in school.* (Final Report United States Office of Education Project No. 2616) Washington, D.C.: United States Government Printing Office, 1966.

TEMPLE, R., & AMEN, E. A study of anxiety reactions in young children by means of a projective technique. *Genetic Psychological Monograph,* 1944, **30,** 59–114.

VELDMAN, D. J., & PECK, R. F. Student teacher characteristics from the pupil's viewpoint. *Journal of Educational Psychology,* 1963, **54,** 346–357.

VELDMAN, D. J., & PECK, R. F. *The pupil observation survey: Teacher behavior from the students' viewpoint.* (Research Methodology Monograph No. 2, Research and Development Center for Teacher Education) Austin: University of Texas Press, 1967.

(Received April 15, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 4, 333–338

EFFECT OF AVAILABILITY ON THE RELATIONSHIP OF MEMORY ABILITIES TO PERFORMANCE IN MULTIPLE-CATEGORY CONCEPT TASKS[1]

DANIEL D. BLAINE AND JACK L. DUNHAM[2]
University of Texas

Six short-term memory tests from Guilford's structure-of-intellect model were administered to 60 subjects. The subjects were then presented with a four-category concept problem in two phases. During the first 36 trials, randomly selected instances were repeated in order to determine subjects' ability to recall previously presented instances. Beginning with the thirty-seventh instance, the previous instance was made available for 30 randomly selected subjects, while the other 30 subjects were not provided with an available instance. Factor analysis of the six tests yielded two factors: Rote Memory and Classes Memory. The ability to recall past instances was significantly correlated with Rote Memory. Presence of an available instance did not significantly facilitate mean performance, but it did significantly alter the relationship between Rote Memory and performance in the concept task.

Recently, investigations concerned with the role of various intellectual abilities in different concept-learning tasks (e.g., Bunderson, 1967; Dunham & Bunderson, 1969; Dunham, Guilford, & Hoepfner, 1968; Lemke, Klausmeier, & Harris, 1967) have established that tests of short-term memory abilities are related to performance in concept-learning tasks. Of specific relevance to this study, Dunham and Bunderson found Associative Memory (French, Ekstrom, & Price, 1963) to have a strong relationship with performance on a multiple-category conjunctive problem administered under standard conditions.

The relationship between memory ability and performance on multiple-category problems can be interpreted as due to the necessity of recalling information from past instances in order to solve such tasks. Overstreet and Dunham (1969) experimentally divided the multiple-category conjunctive problem into two phases: a dimension selection phase and an associative learning phase. In both of these phases the recall of past information seems necessary. In the dimension selection phase, if the subject has been informed of the number of dimensions involved in the solution, he could begin by hypothesizing that a subset of the problem dimensions are the relevant ones. By comparing two instances, a decision could be made about the status of the hypothesis. If two instances were from the same category, and the values of the dimensions involved in the hypothesis were different, then the hypothesis could be rejected. Thus the selection of the relevant dimensions by the use of these rules would necessitate the recall of information from past instances. The same memory requirement is essential in the associative learning phase, since the subject must remember the category labels assigned to the combinations of values determined by the conjunctive rule.

Several investigators have experimentally reduced the necessity of recalling past instances by making various numbers of previously presented instances available. Such availability has consistently been shown to facilitate performance in concept-learning tasks (Blaine, Dunham, & Pyle, 1968; Bourne, Goldstein, & Link, 1964; Pishkin, 1967; Pishkin & Wolfgang, 1965). This result has usually been interpreted as a reduction in the memory requirement of the task.

---

[1] This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Office of Naval Research under Contract N00014-67-A-0126.

[2] Requests for reprints should be sent to Jack L. Dunham, University of Texas, Sutton Hall 319, Austin, Texas 78712.

If availability reduces the memory requirement of the task, the relationship of memory abilities to performance in the task should similarly be altered. That is, the relationship between tests of memory ability and performance in a concept task should decrease when the memory requirement is reduced. Thus, the availability of past instances should not only facilitate performance, but should also reduce the importance of memory abilities in concept-learning performance. This study was designed to test this interaction between memory abilities and availability of past instances. In order to sample a wider range of memory abilities than Dunham and Bunderson, six tests of memory abilities were selected from Guilford's structure-of-intellect model on the basis of their possible relevance to the concept task. In addition, an attempt was made to assess recall of past instances while subjects were solving a multiple-category problem.

METHOD

Subjects

The subjects were 60 undergraduates from introductory educational psychology courses at the University of Texas at Austin.

Materials and Procedures

Tests of six memory abilities from Guilford's (1967; Guilford & Hoepfner, 1966) structure-of-intellect model were administered to all subjects in a group setting prior to their work with the task. The tests dealt with memory for symbolic and semantic units, classes, and implications. The six tests are listed in Table 1.

In the Memory for Symbolic Units test, 15 nonsense trigrams are presented for study. On a test page, 30 such trigrams are listed, and the subjects must determine which of these appeared on the study page. The Memory for Semantic Units test involves the recall of as many meaningful words as possible from 20 words previously studied.

In the Memory for Symbolic Classes test, 10 sets of three trigrams are presented for study. The three trigrams in each set have common attributes, for example, NEC, NEP, NEF, which the subjects must recall in order to recognize which one of four trigrams in several sets of four belong with a set previously studied, for example, NOP, NAR, NER, NUP. The Memory for Semantic Classes test presents sets of three words with common attributes, for example, SILK, WOOL, NYLON, which subjects must recall in order to recognize whether new sets of three words, for example, RAYON, COTTON, FELT, have the same common attributes as sets previously studied.

TABLE 1

MEANS, STANDARD DEVIATIONS, AND RELIABILITIES OF THE SIX MEMORY TESTS AND THE REPEATED INSTANCES MEASURE

Name and identification of test	*M*	*SD*	Reliability[a]
1. Recalled Words (Memory for Semantic Units, MMU)	25.63	4.29	.71
2. Memory for Nonsense Word Classes (Memory for Symbolic Classes, MSC)	13.90	5.13	.83
3. Paired Associates Recall (Memory for Semantic Implications, MMI)	16.45	5.35	.82
4. Classified Information (Memory for Semantic Classes, MMC)	53.80	5.04	.42
5. Number-Letter Associations (Memory for Symbolic Implications, MSI)	13.42	4.95	.73
6. Memory for Listed Nonsense Words (Memory for Symbolic Units, MSU)	51.65	6.06	.65
7. Repeated Instances	6.35	2.20	.71

[a] All reliability estimates are based on the Spearman-Brown estimate of whole-test reliability from separately timed halves except for Repeated Instances for which scores on odd and even items were used.

The Memory for Symbolic Implications test presents a study page of two digit numbers with a letter associated with each. On a test page, the two digit numbers are presented in random order, and subjects must recall the letter associated with each. The Memory for Semantic Implications test is the same type of test with pairs of meaningful words presented on the study page and the first word of each pair presented on the test page. The subjects must recall the word which is associated with each word presented.

All of the concept-learning materials, except for certain segments of the instructions, were presented on IBM 1050 computer terminals under the control of an IBM 1440 system. The concept stimuli were typed out to the subject, and the subject in turn responded by typing in his response at the terminal keyboard.

The task was a four-category, two-dimensional conjunctive concept problem presented by the method of anticipation. The stimuli consisted of the letters A through F, which were always in alphabetical order, but each of the letters could appear in uppercase or lowercase. The task for all subjects was to learn to categorize the instances based upon the uppercase and lowercase combinations of the two relevant letters. The categories were designated 1, 2, 3, and 4.

The concept task was presented in two phases. The first phase consisted of 24 concept instances plus 12 of these 24 which were randomly selected and repeated within the first 24 with either 1, 2, or 3 intervening instances. It was assumed that the subjects more capable of recalling past instances should respond correctly to a greater number of the repeated instances. In this way an attempt was made to assess the subject's ability to recall previously presented instances. In the second phase, the availability conditions were introduced; either zero or one previous instance was available. Of the 60 subjects, 30 were randomly assigned to the zero-availability condition and 30 to the condition which provided one available instance.

In the first phase, all instances were presented under the zero-availability condition. At the end of the first 36 instances, the subjects were branched to one of the availability conditions according to the previous random assignment. In the zero-availability condition, instances were presented such that each instance was spaced under a face-sheet before the next instance was presented. With one instance available, an instance and its correct category remained in view until the next instance had been presented, the subject had responded to it, and had received feedback. The criterion for learning was 20 consecutive correct trials. If the subject had not reached criterion when 76 instances had been presented, the program was terminated.

RESULTS AND DISCUSSION

The means, standard deviations, and reliabilities of the six memory tests and the Repeated Instances test are reported in Table 1. A principal-axis factor analysis of these tests yielded two factors which were rotated to a Varimax criterion. The rotated factor matrix is given in Table 2. Factor I was defined by the two units and the two implications tests, while Factor II was defined by the two classes tests.

The substantial loadings of the units and implications test on the same factor are probably due to both types of tests involving the recall of material which the subjects have previously had a chance to memorize. Both types of tests involve the recall of the same material that was studied. On the other hand, the classes tests require the subjects to recognize whether new items have the same common attributes as any of the sets which they have previously studied. The sets which are studied are made up of different items than those which are presented to the subjects as the test items. Thus, the resultant factor structure could be due to whether the tests involved recalling previously studied material or recalling attributes the studied items had in common in order to deal with new items. On this basis, Factor I was considered as representing a Rote Memory ability, while Factor II represented a Classes Memory ability.

TABLE 2
ROTATED FACTOR MATRIX OF SIX MEMORY TESTS

Test	Factor	
	I	II
1. Recalled Words (MMU)	.82	−.07
2. Memory for Nonsense Word Classes (MSC)	.18	.59
3. Paired-Associates Recall (MMI)	.80	.07
4. Classified Information (MMC)	−.06	.86
5. Number-Letter Association (MSI)	.68	.13
6. Memory for Listed Nonsense Words (MSU)	.56	.33

A *t* test on trials to criterion revealed no significant difference between zero and one past instance available, $t = 1.25$, $df = 58$, $p = .21$. The means for the availability conditions were 62.33 without an available instance and 58.80 with one instance available. Thus, providing a previously presented instance did not significantly facilitate performance in the concept task. This result does not appear to be consistent with the results of previous studies in which availability has been shown to facilitate performance. In the case of Bourne et al. (1964), there is some question as to whether one available instance was effective in facilitating performance in tasks with two relevant dimensions. Further, the instances in the concept task were either positive or negative. In the present study, the task was a multiple-category problem in which the subjects were required to categorize the instances into one of four categories. In the cases where the effect of availability was considered in a four-category problem (Blaine et al., 1968; Pishkin, 1967; Pishkin & Wolfgang, 1965), the minimum level of availability has been one instance per category. This provided the subject with a total number of instances equal to the number of categories in the problem. Thus, in a four-category problem, four instances would be available to the subject. In the present study, only the previous instance was available, regardless of its category designation. This represents a minimum number of available instances and is well below the lower limit of available instances thus far employed in a multiple-category problem. Further, this minimum level of availability was not introduced until after the thirty-sixth instance of the problem had been presented. This would also tend to reduce the effectiveness of the available instance.

To assess the role of memory under the respective ability conditions, two separate regression analyses were performed, with trials to criterion as the dependent variable and factor scores on the Rote Memory and Classes Memory factors as covariables. The factor score regression lines for the two analyses are plotted in Figures 1 and 2. If the memory abilities are required by the multiple-category problem, then the regression lines for the group without an available instance should have negative slopes. The slope of the regression lines in this group for Rote Memory was −7.32 (see Figure 1) and −2.23 for Classes Memory (see Figure 2). The slope of the Rote Memory regression line was significantly different from zero, $F = 5.13$, $df = 1/56$, $p < .05$, while the slope of the Classes Memory regression line was not significant, $F < 1.0$, $df = 1/56$. Thus, high Rote Memory ability was related to successful performance, while the Classes Memory ability was not significantly related to the performance of subjects in the concept task.

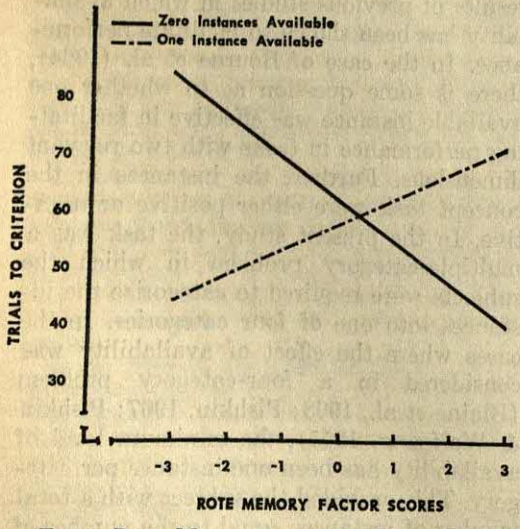


FIG. 1. Rote Memory regression lines for zero and one instance available with trials to criterion as dependent variable.

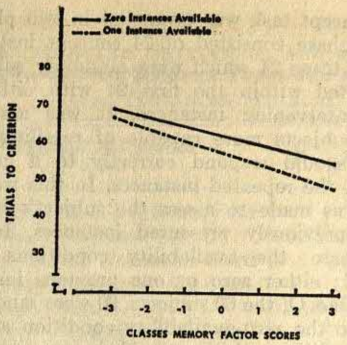


FIG. 2. Classes Memory regression lines for zero and one instance available with trials to criterion as dependent variable.

If the availability of a past instance can be interpreted as a reduction in the memory requirement of the concept task, then a reduction in the relationship between Rote Memory ability and trials to criterion should occur for the group with one past instance available. That is, the memory ability found to be related to performance in the concept task should be less important to performance when an available instance is provided. The slope of the regression line for Rote Memory in the group with one past instance available was 4.49 (see Figure 1). The difference between the slopes of the Rote Memory regression lines for the two availability groups was significant, $F = 7.73$, $df = 1/56$, $p < .01$. Thus, there was a significant difference in the relationship of the Rote Memory ability to performance under the respective availability conditions. The positive slope of the regression line in the group with one past instance available represents not only a reduction in the role of the Rote Memory ability, but a reversal in the character of the relationship. Thus, high Rote Memory ability did not facilitate performance in the group with one past instance available, but rather appeared to be detrimental to performance. However, for this group a comparison of the slope of the regression line with zero only approached significance, $F = 2.65$, $df = 1/56$, $p = .11$.

The positive slope of the Rote Memory regression line with an available instance indicates that the performance of subjects high in Rote Memory ability tends to be inferior to that of subjects low in the ability. If it is assumed that subjects develop strategies consistent with their ability strengths, then this result is not unreasonable. Since the first 36 trials of the concept task were presented without an available instance, subjects high in Rote Memory could develop a strategy capitalizing on their ability to recall past instances. When the available instance was introduced, these subjects continued with the now less effective strategy or they could have changed their strategy and attempted to use the available past instance. The subjects low in the ability to recall past instances would be more likely to develop a strategy less dependent on such an ability. The introduction of an available past instance would less likely be counter to a strategy which relied on some process other than the recall of information from numerous past instances. The available instance could provide low-memory subjects with information which would make their strategies even more efficient than those of subjects capitilizing on their high Rote Memory ability.

The slope of the regression line for the Classes Memory ability was -2.85 in the group with one past instance available. When compared with the regression line for the group with zero instances available, the difference was not significant, $F < 1.0$, $df = 1/56$, nor was this slope significantly different from zero, $F < 1.0$, $df = 1/56$. Thus, the Classes Memory ability was not related to performance in the concept task, and its relationship to performance was not altered by the introduction of an available past instance. On the other hand, high Rote Memory ability was related to successful performance in the concept task, and this relationship was significantly altered by the introduction of an available past instance.

In this study it was assumed that it would be necessary to recall information from past instances in order to solve multiple-category problems. One way in which the subjects could recall such information is by recalling both the instance and its correct response. The first phase of the task in which randomly selected instances were repeated was an attempt not only to assess the subjects' ability to recall previously presented instances, but to assess the ability from within the task in which the ability would be related to performance. However, since 13 of the 60 subjects were in a criterion run within the first 36 instances, the analyses concerning the Repeated Instances test includes only those subjects who did not solve until after the first phase had been completed and the availability conditions had been introduced.

In an attempt to evaluate the nature of the Repeated Instances test, it was correlated with the factor scores on the Rote and Classes Memory factors. The correlations between Repeated Instances and Rote and Classes Memory were .29 ($p < .05$) and $-.02$, respectively. This indicates that the ability to recall previously presented instances is most closely associated with the Rote Memory ability. This result is not surprising in view of the nature of the concept task. In order to respond correctly to a greater number of repeated instances, the subjects would have to recall an instance and its correct feedback. Thus the Repeated Instances test would require subjects to recall material which he had previously studied. This type of memory is similar to the type of memory identified by the Rote Memory factor. Within the zero-availability group, the Repeated Instances test correlated $-.32$ with trials to criterion indicating a positive relationship between the ability to recall previously presented instances and performance in the concept task. In the group with one available instance, the correlation between Repeated Instances and trials to criterion was $-.06$. Thus the ability to recall previously presented instances appeared to be more important in the group with zero past instances available. The slopes of the regression lines with Repeated Instances as a covariable predicting trials to criterion for the groups with zero and one instance available were -1.61 and $-.55$, respectively. However, neither of these slopes was significantly different from a zero slope, $F = 1.45$, $df = 1/43$, $p = .23$, and $F < 1.0$, $df = 1/43$.

In conclusion, it was found that the Rote Memory ability was significantly related to performance in the concept task while a Classes Memory ability was not. The introduction of an available past instance significantly altered the relationship between Rote Memory and performance even though it did not effect significant differences in mean performance. However, since the regression lines for Rote Memory intersected well within the range of the factor scores, the introduction of the available past instance cannot be interpreted as simply as removal or reduction of the memory requirement.

REFERENCES

BLAINE, D. D., DUNHAM, J. L., & PYLE, T. W. Type and amount of available past instances in concept learning. *Psychonomic Science,* 1968, **12,** 159–160.

BOURNE, L. E., JR., GOLDSTEIN, S., & LINK, W. E. Concept learning as a function of availability of previously presented information. *Journal of Experimental Psychology,* 1964, **67,** 439–448.

BUNDERSON, C. V. *Transfer of mental abilities at different stages of practice in the solution of concept problems.* (Research Bulletin RB-67-20) Princeton, N. J.: Educational Testing Service, 1967.

DUNHAM, J. L., & BUNDERSON, C. V. The effect of decision-rule instruction upon the relationship of cognitive abilities to performance in multiple-category concept problems. *Journal of Educational Psychology,* 1969, **60,** 121–125.

DUNHAM, J. L., GUILFORD, J. P., & HOEPFNER, R. Multivariate approaches to discovering the intellectual components of concept learning. *Psychological Review,* 1968, **75,** 206–221.

FRENCH, J. W., EKSTROM, R. B., & PRICE, L. A. *Manual for Kit of Reference Tests for Cognitive Factors.* Princeton, N. J.: Educational Testing Service, 1963.

GUILFORD, J. P. *The nature of human intelligence.* New York: McGraw-Hill, 1967.

GUILFORD, J. P., & HOEPFNER, R. Structure-of-intellect tests and factors. *Reports from the psychological laboratory.* (No. 36) Los Angeles: University of Southern California, 1966.

LEMKE, E. A., KLAUSMEIER, H. J., & HARRIS, C. W. Relationship of selected cognitive abilities to concept attainment and information processing. *Journal of Educational Psychology,* 1967, **58,** 27–35.

OVERSTREET, J. D., & DUNHAM, J. L. Effect of number of values and irrelevant dimensions on dimension selection and associative learning in a multiple-concept problem. *Journal of Experimental Psychology,* 1969, **79,** 265–268.

PISHKIN, V. Availability of feedback-corrected error instances in concept learning. *Journal of Experimental Psychology,* 1967, **73,** 318–319.

PISHKIN, V., & WOLFGANG, A. Number and type of available instances in concept learning. *Journal of Experimental Psychology,* 1965, **69,** 5–8.

(Received April 29, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 4, 339-344

EFFECTS OF VERBAL AND PICTORIAL ELABORATIONS ON ASSOCIATIVE LEARNING AND RESPONSE LEARNING IN A CHILDREN'S PAIRED-ASSOCIATE TASK[1]

STEVE LYNCH[2]
University of California, Davis

AND

WILLIAM D. ROHWER, JR.
University of California, Berkeley

Two experiments were conducted to test the hypotheses that (*a*) sentential elaborations of materials would affect only the associative learning phase of paired-associate learning, and (*b*) pictorial elaborations would affect only the response learning phase of paired-associate learning. It was assumed that the matching task allows the assessment of associative learning independent of the effects of response learning. Results of the first study confirmed the first hypothesis but not the second. In the second study, the first hypothesis was again supported, and pictorial elaboration was found to facilitate both associative learning and response learning.

It has been some time since serious doubt could be raised concerning the effectiveness of pictorial and verbal elaboration in promoting paired-associate learning efficiency. In a variety of studies, the amount learned has been considerably increased by presenting noun pairs in the context of simple declarative sentences (Rohwer & Lynch, 1966, 1967; Rohwer, Lynch, Levin, & Suzuki, 1967; Rohwer, Lynch, Suzki, & Levin, 1967), or, by presenting pictures of the objects named by the nouns (Rohwer, Lynch, Levin, and Suzuki, 1967). While the instructional implications of such findings are relatively clear, little progress has been made in explaining the processes through which these elaborations facilitate learning. Explanations of the effectiveness of sentence elaboration in terms of lower intralist similarity (Rohwer & Lynch, 1967), semantic constraint (Rohwer & Lynch, 1966), and the availability of sentence contexts at the time of information retrieval (Rohwer, Shuell, & Levin, 1967) have all been tested and found wanting.

The present research is an attempt to relate pictorial and verbal elaborations to subject processes through the utilization of a methodology proposed and verified by Underwood, Runquist, and Shulz (1959). The paired-associate task may be divided conceptually into two subprocesses: associative learning and response learning. The subject first acquires the individual response and then learns the association between the response and the stimulus. Were this true, one might speculate that variations in the contexts of noun-pair presentation would affect the two subprocesses differently. For the present work, it was specifically hypothesized that sentences would affect only the associative learning phase and pictures only the response learning phase of paired-associate learning.

EXPERIMENT I

In order to isolate the associative learning and response learning phases of the paired-associate task, it is necessary to depart from standard methodology. In the present study, associative learning was measured by a matching task in which all stimuli and responses were presented to the subject at the time of the test trial. Under such conditions, the necessity for recall of individual items is eliminated, and the number of correctly matched pairs is assumed to represent the number of correct associations independent of the availability of the response items in the various experimental groups. The measure of response

---

[1] This research was supported, in part, by funds received through the University of California Academic Senate Committee on Research (3-54321-3-19900-5) and by a grant with the National Institutes of Health (HD-03869). An abbreviated version of this paper was presented at the annual meeting of the American Educational Research Association at Minneapolis, March 1970.

[2] Requests for reprints should be sent to Steve Lynch, Department of Education, College of Letters and Science, University of California, Davis, California 95616.

learning was simply the number of responses correctly recalled.

Method

Materials and design. The principal factors in a 2 × 2 × 2 design were mode (printed words versus pictures), verbalization (naming versus sentence), and sequence (associative learning tested first versus response learning tested first). All training pairs were presented by slide projection. Half the subjects saw printed word pairs (for example, COW-TIE), and half the subjects saw line drawings of the objects named by the words. In each of these groups, half of the subjects heard the experimenter say the names of the objects (or, heard the experimenter read the words), and half heard the experimenter utter a simple declarative sentence in which the noun pair was embedded (e.g., "the cow squeezes the tie"). A list of the 20 pairs of common nouns used appears in Table 1.

TABLE 1
PAIRING-TRIAL MATERIALS USED IN EXPERIMENTS I AND II

Naming		Sentence
Top	Moon	The top uses the moon.
Chair	Shoe	The chair hurts the shoe.
Picture	Mailbox	The picture fixes the mailbox.
Ring	Lamp	The ring hits the lamp.
Cow	Tie	The cow squeezes the tie.
Key	Pillow	The key climbs the pillow.
Iron	Candle	The iron rubs the candle.
Ball	Feather	The ball covers the feather.
Pencil	Cake	The pencil cuts the cake.
Brush	Skate	The brush tickles the skate.
Hat	Tire	The hat rolls the tire.
Bird	Dress	The bird taps the dress.
Ax	Kite	The ax sweeps the kite.
Pipe	Fish	The pipe smokes the fish.
Butterfly	Whistle	The butterfly chases the whistle.
Ship[a]	Buggy	The ship moves the buggy.
Money[a]	Safetypin	The money holds the safetypin.
Rope[a]	Jar	The rope opens the jar.
Book[a]	Clock	The book breaks the clock.
Chain[a]	Bowl	The chain fills the bowl.

[a] These items were experimental items in Experiment I and distractor items in Experiment II.

In order to test the subjects on both associative and response learning, the sequence of test tasks was reversed for half the subjects in each condition. That is, after training, half the subjects were asked to perform the matching task first and then, immediately, the response recall task, and half the subjects responded in the reverse order of tasks. For the matching task, each subject was given a booklet arranged so that when it was opened, the stimuli appeared on the left-hand page and the responses on the right-hand page. Each response was numbered, and the subjects task was to write the response number next to the correct stimulus term. All stimuli and responses appeared in different orders on the two test trials, and these were different from the orders of study trial presentation. In the same booklet, an additional page was provided on which the subject recorded all the responses he could recall.

Subjects. The sample consisted of the entire sixth-grade population in a local public school servicing a middle-class residential neighborhood.

Procedures. Subjects were drawn at random and assigned at random to experimental groups of size seven. The subjects were seated at tables in the experimental room, given test booklets, and instructed that they were to learn the pairs in such a way as to be able to remember and match the pairs of words (pictures) which they would be seeing on the screen. Two examples were given orally and the experimenter answered any questions in an effort to assure that all subjects understood the task. Examples included instruction in the nature of the matching task, but no mention was made of the response recall task. Pairs were presented at a 4-second rate, and as each pair appeared on the screen, the experimenter uttered the appropriate verbalization. Immediately following the study trial, the subjects were instructed to open their booklets and told to begin the matching task (Sequence 1) or to write down all the responses they could recall (Sequence 2). For each sequence, completion of the first task was immediately followed by instructions to begin the second task. Two minutes were allowed for the completion of each task.

Results[3]

Since independent univariate analyses of variance were performed on both the measures of learning, a stringent criterion for significance was adoped; only those sources of variance significant at the .01 level or better were accepted as indicative of nonrandom mean difference. The means for the results of Experiment I appear in Table 2.

Associative learning. Since sequence was not a variable of interest, both mode and verbalization were nested within it. Verbalization within sequence was a significant source of variance ($F = 14.98$, $df = 2/48$, $p < .01$) such that sentences produced more correct associations than did the naming control in both sequences. In addition, the effect of mode within verbalization within sequence ($F = 10.84$, $df = 2/48$, $p < .01$) was significant. When this effect was partitioned into the sums of squares associated with mode within each level of verbalization in each sequence, pictures were found to facilitate associative learning only when presented in conjunction with sentences. That is, pictures without sentences, as predicted, did not facilitate associative learning.

---

[3] All post hoc comparisons were made by the Scheffé method with $p = .01$.

Response learning. Mode within verbalization within sequence was a significant source of variance ($F = 4.08$, $df = 4/48$, $p < .01$). A breakdown of the variance into components associated with mode within each level of verbalization and sequence revealed, however, that pictures facilitated response learning only in the sentence condition of Sequence 2. That is, when sentences are verbalized with printed words or pictures, and response learning is tested before associative learning, the sentence/picture condition is associated with better response learning than the sentence/printed word condition. Verbalization within sequence, as predicted, did not facilitate response learning. ($F = 1.33$, $df = 2/48$, $p > .05$).

The tests of interest, then, were found to be in support of the original hypotheses with respect to associative learning, but not response learning. The finding that sentences with pictures facilitated response learning was regarded as suggestive but not supportive of the hypothesis concerning response learning. Two factors suggested that an additional experiment was needed in order to adequately test the hypotheses associated with the two-stage analysis. First, as has been mentioned, the result for sentences with pictures, while not convincing, was in the predicted direction. Second, it was realized that the test tasks required word output (matching of words or written recall) regardless of the nature of the study trial materials. It seemed possible that training subjects with pictures and subsequently testing them with word output might be deleterious to learning. The observed results might then be an artifact of method. Accordingly, a second study was designed to explore the latter possibility and retest the original hypotheses.

TABLE 2

MEAN NUMBERS OF CORRECT RESPONSES IN EXPERIMENT I AS A FUNCTION OF MODE, VERBALIZATION, AND SEQUENCE

Mode	Sequence 1		Sequence 2	
	Naming	Sentence	Naming	Sentence
Associative learning				
Words	6.43	9.64	6.71	7.79
Pictures	6.71	13.86	6.71	13.71
Response learning				
Words	8.14	7.57	6.07	4.86
Pictures	6.07	7.64	6.43	9.79

Note.—$M_{SE} = 20.15$, $df = 48$ for associative learning; $M_{SE} = 7.37$, $df = 48$ for response learning.

EXPERIMENT II

In order to allow for congruent input–output modes it was necessary to make two changes in experimental design and procedure. First, since one could not ask the subject to draw his responses, a recognition task was used for both learning measures. Second, the recognition task must include more items than the subject originally learned. That is, some proportion of extralist distractor items was necessary to avoid the possibility that the subject might claim to recognize all items. Two other changes were also instituted, namely, that the subjects were instructed in the nature of the response learning task, and that they were instructed to match pairs and, at the same time, to indicate which items were recognized. That is, response learning and associative learning were measured simultaneously rather than separately on successive tasks.

Method

Materials and design. The stimulus materials were those used in the previous experiment, except that the study trial was made up of 15 pairs rather than 20. The remaining pairs appeared in the recognition task as distractor items. Measures of associative learning and response learning were obtained simultaneously. Each subject was given a test booklet in which he was required to circle all recognized items and indicate which items belonged in pairs. The test booklets were arranged in such a way that when the subject opened his booklet, all stimulus and response items, plus 10 distractor items, were displayed before him on two pages. The placement of items was randomly arranged and a different random order was used on each of the two test trials. Each word (or picture) was numbered from 1 through 40. At the edge of the second page, lines were provided on which the subject was to record the pairs of numbers which went with each pair presented during the study trials.

The principal factors in a 2 × 2 × 2 design were congruence (congruent input–output modes versus incongruent input–output modes), mode (printed words versus pictures), and verbalization (naming versus sentences). Materials, both printed and pictorial, were presented by slide projection. In congruent conditions, subjects who viewed pictures on the study trial responded to pictures on the recognition test trials, and subjects who viewed printed words responded to printed words. In the incongruent conditions, picture training was followed by printed-word recognition, and vice-versa. The mode and verbalization factors were as described in Experiment I.

Procedures. Subjects were run in groups of size 12. Test booklets were passed out to the subjects and instructions given in the procedures to be used. Two examples were given visually to be certain that all subjects understood the instructions. During study trials, paired-associate items were presented at a 4-second rate. At the end of each study trial, the subjects were instructed to open booklets and begin. Three and one-half minutes were provided for each test trial.

Subjects. The sample of 96 sixth-grade subjects were drawn from a Northern California public school district in a city different from the city utilized in the first experiment. The social class of the neighborhood serviced by the two districts were similar, however, and the samples were considered comparable. Subjects were drawn at random and were randomly assigned to experimental groups.

Results[4]

Associative learning. Since independent univariate analyses of variance were performed on the two dependent measures, only those sources of variance significant at the .01 level or better were accepted as indicative of nonrandom mean differences. The main effects of congruence ($F = 14.59$, $df = 1/88, p < .01$) and verbalization ($F = 18.22$, $df = 1/88, p < .01$) were both significant sources of variance. Congruent input–output modes produced greater associative learning than incongruent input–output modes and sentence conditions were associated with better learning than naming conditions. An examination of the significant interaction of modes and congruence ($F = 9.06$, $df = 1/88$, $p < .01$) revealed that pictures significantly facilitated associative learning only when the input–output modes were congruent. The means for these relationships appear in Table 3. No other between-groups relationships were significant.

TABLE 3

MEAN NUMBERS OF CORRECT RESPONSES IN EXPERIMENT II AS A FUNCTION OF CONGRUENCE, MODE, AND VERBALIZATION

Mode	Congruent		Incongruent	
	Naming	Sentence	Naming	Sentence
Associative learning				
Words	3.71	5.54	3.29	5.21
Pictures	6.08	8.25	3.00	5.00
Response learning				
Words	11.46	10.50	7.08	9.17
Pictures	12.50	13.21	12.71	13.13

Note.—$M_{SE} = 10.32$, $df = 88$ for associative learning; $M_{SE} = 10.84$, $df = 88$ for response learning.

Response learning. Since the subject was required to circle all stimuli and responses which he recognized, there was a choice of dependent variables available for analysis. The results for both stimulus learning and response learning were analyzed and found to be identical. In order to remain consistent with Experiment I, however, only the results for response learning will be presented here. Congruence ($F = 8.63$, $df = 1/88, p < .01$) and mode ($F = 49.21$, $df = 1/88$, $p < .01$) were significant sources of variance, congruent input–output modes and pictures both leading to efficient response learning. A significant interaction of congruence and mode was observed ($F = 9.42$, $df = 1/88$, $p < .01$) such that pictures were unaffected by incongruence but the shift from learning printed words to recognizing pictures had a rather drastic negative affect on learning efficiency. Verbalization was not a significant source of variance ($F = 1.40$, $df = 1/88$, $p > .25$) indicating that sentence elaborations have no facilitating effect on the response learning phase of paired-associate learning.

---

[4] All post hoc comparisons were made by the Scheffé method with $p = .01$.

DISCUSSION

There are two difficulties in interpreting the results of the present study. The first of these is concerned with an inconsistency in the results of the two experiments. It may be recalled that the principal reason for engaging in the second experiment was the possibility that input–output incongruence had depressed response learning for pictures in the first experiment. In the second experiment, pictures led to the predicted facilitation of response learning, but incongruence did not depress the learning. The most likely source of this inconsistency is in the instructional change which, in Experiment II, directed the subjects' attention to the fact that they would be tested for both associative and response learning, hence possibly inducing a learning strategy difference between the two studies. It is important to note that in the typical paired-associate task, the subject's attention is not directed toward the response learning aspects of the task. The present results then, while suggestive, may not be directly relevant to interpretations of outcomes in the typical paired-associate task insofar as those interpretations concern the effects of response learning on the whole task.

The second difficulty arises from the use of distractor items in the matching task for associative learning. It may be recalled that the matching procedure was intended to eliminate the necessity for recall of individual items, thus equating the different experimental groups on response learning. The inclusion of the 10 extra-list distractor items introduced an unwanted process into the task, however. Clearly, before the subject can perform the associative task he must discriminate between extra-list distractor items and experimental items. Since pictures facilitate response learning, it is likely that they will also facilitate response discrimination in a recognition task, thus giving picture condition subjects an advantage in associative learning which is unrelated to the associative subprocess. It would appear that an additional experiment involving independent response and associative learning groups will be necessary to investigate this possibility.

In spite of the difficulties mentioned above, the initial hypotheses concerning the effect of sentences on associative learning and pictures on response learning seem to be adequately supported. Sentences were observed to facilitate only associative learning. Pictures facilitated both response learning and associative learning, although the latter result is subject to some qualification.

The method of two-stage analysis has, in addition, several intriguing possibilities. First, it may ultimately direct researchers toward more accurate speculation about the nature of those covert subject processes which cannot be measured directly. For example, researchers currently working in the area of mental imagery (Milgram, 1967; Paivio & Yarmey, 1966; Rohwer, Lynch, Levin, & Suzuki, 1967; Rohwer, Lynch, Suzuki, & Levin, 1967) have assumed that the measurement of learning by verbal or written output is a satisfactory procedure. Given the present results concerning the effects of incongruent input–output modes, the inferences drawn from previous studies on the nature of subject processes and learning efficiency must be held in question. The translation of mental images into verbal output, whether the locus of the effect is at the time of storage or at the time of recall, is clearly an issue requiring further investigation. Second, the method should allow the classification of elaborative variables into broad categories depending upon the functional relationship of the variables to the subprocesses. For example, Rohwer, Lynch, Levin, and Suzuki (1967) have shown that action pictures are better facilitators of paired-associate learning than pictures which depict a nonaction, locational relationship, and that the latter are more effective than nonaction, nonlocational pictures. It may be that this ordered effect is not unidimensional with respect to the two paired-associate subprocesses. The method could also profitably be applied to the auditory and visual modes separately, investigating the effects of audition alone and of pictures or words without audition.

Finally, other researchers (Milgram, 1967; Yuille & Paivio, 1968) have shown that instructions to subjects to create their own mental images or their own sentence contexts will facilitate paired-associate learning. These instructional variations have been treated as though they were analogous to the materials variations in which sentences or pictures are provided to the subject. The two-stage analysis may aid in deciding whether or not subject processes are the same in the two experimental procedures.

REFERENCES

MILGRAM, N. A. Retention of mediation set in paired-associate learning of normal children and retardates. *Journal of Experimental Child Psychology*, 1967, **5**, 341–349.

PAIVIO, A., & YARMEY, A. D. Pictures versus words as stimuli and responses in paired-associate learning. *Psychonomic Science*, 1966, **5**, 235–236.

ROHWER, W. D., JR., & LYNCH, S. Semantic constraint in paired-associate learning. *Journal of Educational Psychology*, 1966, **57**, 271–278.

ROHWER, W. D., JR., & LYNCH, S. Form class and intralist similarity in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 551–54.

ROHWER, W. D., JR., LYNCH, S., LEVIN, J. R., & SUZUKI, N. Pictorial and verbal factors in the efficient learning of paired associates. *Journal of Educational Psychology*, 1967, **58**, 278–284.

ROHWER, W. D., JR., LYNCH, S., SUZUKI, N., & LEVIN, J. R. Verbal and pictorial facilitation of paired-associate learning. *Journal of Experimental Child Psychology*, 1967, **5**, 294–302.

ROHWER, W. D., JR., SHUELL, T. J., & LEVIN, J. R. Context effects in the initial storage and retrieval of noun pairs. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 796–801.

UNDERWOOD, B. J., RUNQUIST, W. N., & SHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *Journal of Experimental Psychology*, 1959, **58**, 70–78.

YUILLE, J. C., & PAIVIO, A. *Imagery and verbal mediation instructions in paired-associate learning*. (Research Bulletin No. 69) London, Canada: University of Western Ontario, Department of Psychology, 1968.

(Received May 4, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 4, 345–351

RELATIONSHIPS OF TEACHING METHOD, SOCIOECONOMIC STATUS, AND INTELLIGENCE IN CONCEPT FORMATION

JOHN T. GUTHRIE[1]
Johns Hopkins University

A 4 × 2 × 2 design, including four instructional treatments and two levels of socioeconomic status (SES), and two levels of IQ, was used to study methods for facilitating the formation of a one-dimensional language concept. The criterion concept-formation task required the oral production of sentences which contained a novel grammatical concept. Sixty-four Negro fifth and sixth graders were run individually. The results indicated that (*a*) concept formation was facilitated by the same instructional procedures, the presentation of rules and the application of rules, for both high and low-SES subjects; and (*b*) that the high-SES group learned in significantly fewer trials than the low-SES group.

A recurrent theme in contemporary discussions of learning and instruction is that teaching procedures should be tailored to the cognitive capabilities of the learner. It is reasoned that since different individuals manifest different profiles of learning ability, the success of a given instructional strategy depends on the extent to which it is accommodated to the intellectual characteristics of the student. For example, Jensen (1969) asserts that "educational researchers must discover and devise teaching methods that capitalize on existing abilities for the acquisition of those basic skills which students will need [p. 117]." In addition Cronbach (1969) urges that "new kinds of instruction be developed to fit diverse patterns of ability [p. 347]."

The primary issue to which this study addresses itself is whether different forms of instruction interact with socioeconomic status (SES) and intelligence in the facilitation of concept attainment. Pertinent to this issue is the fact that Jensen (1968) has shown that SES is correlated .35–.40 with IQ, the high-SES children having higher IQs than their low-SES peers. On the contrary, the two SES groups do not differ on basic associative learning ability according to Jensen (1969). The basic associative learning ability is measured by simple learning tasks such as paired-associate or serial learning.

One possible implication of Jensen's finding is that middle-SES children should be taught with methods that capitilize on the well-developed abstract verbal ability of these children. In turn, lower-SES children should be taught with methods which do not rely on such abstract verbal functioning, but which depend primarily on basic associative learning ability. On the other hand, it is possible that abstract verbal functioning is necessary for all types of instruction, and any instructional procedure which attempts to avoid reliance on this ability may be ineffective. Thus, empirical research is required to investigate the relationships of the factors of teaching method, SES, and IQ.

The second purpose of the study was to investigate further the effect of the presentation of rules on concept formation. In a previous experiment (Guthrie & Baldwin, 1970), it was found that subjects who were taught to verbalize a rule which governed a grammatical concept did not learn to generate instances of the concept any faster than subjects who were not taught the rule. However, the concept was simple, involving only the production of a two-word response to a one-word stimulus; that is, the subjects were required to say "an orange" when presented with the word "orange," and to respond "a grape" when presented the word "grape," etc. In the presence of vowel-initial words, subjects were required to say "an" before the word and in the presence of consonant-initial words, subjects were required to say "a" before the word. In the present study, the same grammatical concept was used except that the concept was embedded in a sentence. In other words, the criterion concept-formation task required the subjects to generate a response consisting of a complete sentence in the presence of a stimulus sentence. Thus the question is raised of whether rules will facilitate concept formation with this more difficult, complex, and meaningful task, although they failed to affect learning on the simpler, one-word task.

---

[1] Requests for reprints should be sent to John T. Guthrie, Director, Special Education, John F. Kennedy Institute, Baltimore, Maryland 21205.

METHOD

Subjects

The original sample consisted of 83 Negro children drawn from the fifth- and sixth-grade classes of three schools. Of this original group, 19 subjects were dropped. Fifteen subjects failed a pretest which consisted of a vowel-consonant discrimination task. Since all subjects were required to pass this pretest before receiving the instructional treatments, the subjects in all treatment conditions were homogeneous with respect to this stimulus-discrimination skill which was basic to the criterion concept-formation task. Thus the attrition did not produce bias toward a particular treatment; rather it produced homogeneity among the groups. Finally, four subjects were eliminated at random from several of the treatment conditions to create equal numbers of subjects in each of the treatment conditions in order to facilitate the analysis of variance. The final sample of 64 subjects had a mean Kuhlman-Anderson IQ of 103.23 with a standard deviation of 10.46; the median IQ was 104.50.

The SES of the subjects was determined with the Hollingshead occupational scale (Hollingshead & Redlich, 1958). The seven levels of the scale include (*a*) executives and proprietors of large concerns and major professionals; (*b*) managers and proprietors of medium-sized businesses and lesser professionals; (*c*) administrative personnel of large concerns, owners of small independent businesses, and semiprofessionals; (*d*) owners of little businesses, clerical and sales workers, and technicians; (*e*) skilled workers; (*f*) semiskilled workers; and (*g*) unskilled workers. The subjects who had one or more breadwinner in the family at the level in *d* above, or higher, were classified as high SES, whereas other subjects were designated as low SES. The experimental design was a 4 × 2 × 2 factorial including four treatment conditions, two levels of IQ, and two levels of SES. The two levels of IQ were represented by subjects above the median of 104.50 (high IQ) and below the median (low IQ). The two levels of SES were the high and low groups as defined above with the Hollingshead scale. These variables were crossed and four cells were established. From each cell, the subjects were randomly assigned to one of the four treatment conditions.

TABLE 1

COMBINATION OF EXPERIMENTAL TASKS TO FORM TREATMENT CONDITIONS

Experimental tasks	Treatment conditions			
	I	II	III	IV
Pretest	1	1	1	1
Instructional tasks	2 3a 4	2 3b	2	
Concept formation	5	5	5	5

Note.—The numbers refer to the tasks as follows: 1 = vowel–consonant discrimination, 2 = rule verbalization, 3 = rule application (a = appropriate for production of instances; b = appropriate for concept formation), 4 = production of instances, 5 = concept formation. The tasks were administered in the sequence shown in the columns.

Treatment Conditions

The treatment conditions were constructed by combining the experimental tasks systematically (see Table 1). Treatment Condition I consisted of all of the following experimental tasks administered successively: vowel–consonant discrimination (Task 1), rule verbalization (Task 2), rule application appropriate for the production of instances (Task 3a), production of instances (Task 4), and concept formation (Task 5). Treatment Condition II was the same as Treatment I except that the production of instances (Task 4) was omitted. This change from Treatment I necessitated a modification of the rule application task to render it appropriate for the concept-formation task (Task 5) rather than production of instances (Task 4). Treatment Condition III was identical to Treatment II except that the rule application task was excluded. Thus, Treatment III included Tasks 1, 2, and 5. Treatment IV contained Tasks 1 and 5, differing from Treatment III only by the absence of Task 2.

Experimental Tasks

The words used in each task were drawn from a list of 172 vowel-initial and 108 consonant-initial nouns or noun phrases. The words were drawn randomly without replacement from the list for each task.

1. *Vowel–consonant discrimination.* English words were presented to the subjects successively with a tape recorder. A total of 40 words was arranged in blocks of four. Each block included two consonant-initial words and two vowel-initial words occurring in random order. The subjects labeled each word as beginning with a vowel or consonant, and the experimenter provided feedback of "yes" or "no" following each response. The task was discontinued following the attainment of

40 trials. Any subject who failed to attain criterion was replaced by another subject selected at random.

2. *Rule verbalization.* The following rule was read to the subject by the experimenter: "After the word *the,* K comes before consonants and T comes before vowels." Next the first nine words of the rule were read by the experimenter and the subject completed the statement of the rule aloud. The portion of the rule read by the experimenter was reduced, and the portion of the rule verbalized by the subject was increased until the subject verbalized the entire rule twice correctly.

3. *Rule application.* The application task was initiated by presenting a stimulus to the subject. The stimulus was either a single critical word or a sentence and a critical word which was to be substituted for another word in the sentence. The first step was to ask the subject whether the critical word began with a vowel or a consonant. Second, the subject was required to verbalize the portion of the rule which assigned a response, that is, K or T, to words with the type of sound identified in the first step. Third, the stimulus was presented again and the subject was requested to produce a composite oral response which included both the critical word and the appropriate consonant. When the application task was intended to make the rule applicable to the production of instances (Task 4), the stimulus was a single critical word; when the application task was intended to make the rule applicable to concept formation (Task 5), the stimulus was a sentence with instructions to substitute the given critical word for a word in the sentence. The stimuli used in the two forms of the application task were the same as those used in Tasks 4 and 5. An example of the rule application procedure when the rule is applied to a single critical word is as follows:

> Experimenter: Apple. Does apple begin with a vowel or consonant?
> Subject: A vowel.
> Experimenter: What comes before words that begin with vowels?
> Subject: T.
> Experimenter: If I said apple, you would say ______.
> Subject: T apple.

All of the steps in the application procedure were administered until the subject made correct responses at Step 3 on three successive trials. Then Step 2 was omitted from the procedure until the subject attained correct responses at Step 3 for three successive trials. Next, Steps 1 and 2 were omitted and Step 3 was presented alone for one trial. At this point the stimuli were presented successively until the subject attained criterion or completed all of the trials in the task.

Throughout the administration of this task, feedback consisting of "No, try again," was provided following errors on Steps 1 or 2. No feedback was given following correct responses. The feedback given for the third step was "yes" following correct responses and "no" plus the correct response following errors.

4. *Production of instances.* Words, which included only nouns or noun phrases, were presented one at a time to the subjects with a tape recorder, and the subjects responded by saying "K" or "T" and saying the word. For example, given the word "apron," the subject responded "T apron;" and given "fireman," the subject said "K fireman." A total of 60 words was available, half of them vowel-initial and half consonant-initial words. In every block of three words, at least one word of each type was included. Feedback, consisting of "yes" following correct responses and "no" following incorrect responses, was provided on every trial. The subjects were run to a criterion of 12 consecutive correct responses or to the limit of 60 trials. All subjects who were given this task were required to attain criterion before progressing to the next task. One subject failed to reach criterion before the limit of 60 trials and was replaced by another subject at random.

5. *Concept formation.* On each trial in this task, a sentence was presented which contained the letter K or T preceding a noun or noun phrase. Following the presentation of the sentence, the subjects were instructed to substitute a given noun for the noun in the sentence which followed the consonant K or T. The subject responded by saying the sentence with the new word preceded by the appropriate consonant. For example, the subject was given the following stimulus: "That girl was the K girl scout. Change girl scout to outlaw." The subject was then expected to respond by saying: "That girl was the T outlaw." After each response, the subject was given feedback consisting of "yes" following correct responses, and "no" plus the presentation of the correct response following errors.

A total of 72 sentences was available for tape-recorded presentation. The sentences were arranged in blocks of six, each block including an equal number of vowel-initial and consonant-initial words as the critical words to be substituted into the sentence. In addition, four of the sentences in each block contained a word in the sentence which had a different type of initial phoneme (i.e., vowels versus consonants) than the critical word. For these sentences, the subject was required to include in his response a consonant, K or T, which was different from that presented in the original stimulus sentence. Concomitantly, two of the sentences in each block contained a word with the same type of initial phoneme as the critical word. On these sentences, the correct response required the same consonant (i.e., K or T) as presented in the stimulus sentence. Within each block the occurrences of vowel-initial and consonant-initial critical words and the requirement for the same or different type of consonant in the subject's response were determined at random. The criterion for the task was correct responses on all trials in two consecutive blocks of sentences. This criterion insured that all subjects responded correctly on six sentences with consonant-initial critical words and six sentences with a vowel-initial critical word. The criterion also required the production of eight sentences in which the single consonant preceding the critical word was different from the consonant in the stimulus sentence and the production of four sentences where the consonant was the same as that of the stimulus sentence.

In brief, the treatment conditions represent different types of instructional strategies for facilitating concept formation. All of the treatments are preceded by a pretest and followed by the criterion concept-formation task. The different instructional strategies vary in the extent to which the rule is provided; the rule is absent in Treatment IV and it is presented without application in Treatment III. The strategies also vary in the degree of application of the rule: in Treatments I and II, application training is provided whereas it is excluded from Treatment III.

The criterion concept-formation task, which requires the placement of K or T before consonant-initial words and vowel-initial words, was selected to render the results comparable to those of the previous experiment (Guthrie & Baldwin, 1970). This task is similar to the task involving the indefinite article "a" and "an" used previously.

Procedure

The experimenters worked individually with each subject in a small room in the school the subject attended. Both the experimenter and the subject sat at a table on which a Uher tape recorder was placed. All of the materials for Tasks 1, 4, and 5 were presented with a tape recorder. The materials for Tasks 2 and 3 were read by the experimenter since some individual monitoring of the task was needed for each subject. On Tasks 1, 4, and 5, the experimenter maintained a written record of each response as correct or incorrect. The time spent with each subject was approximately 35 minutes.

RESULTS

The trials to criterion on the concept-formation task were analyzed with a 4 × 2 × 2 analysis of variance. The analysis included four treatment conditions, two levels of student IQ, and two levels of SES. The limit in the concept-formation task was 72 trials. However, no subject who attained criterion on the task required more than 40 trials to meet the criterion. Consequently, the nonlearner contributed an extraordinary amount of error variance to the data. To reduce this source of error variance, the nonlearners were assigned a score of 50 rather than their original raw score of 72. These transformed data were then analyzed with the 4 × 2 × 2 analysis of variance. The results were that significant main effects were observed for treatment conditions ($F = 23.15$, $df = 3/48$, $p < .001$) and SES ($F = 4.95$, $df = 1/48$, $p < .05$). Neither the main effect for IQ nor any of the interaction effects were significant. See Table 2 for the group means and standard deviations on which this analysis of variance was based.

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR TRIALS TO CRITERION ON CONCEPT FORMATION TASK

Treatment conditions	High IQ				Low IQ			
	High SES		Low SES		High SES		Low SES	
	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
I	1.25	.44	7.25	8.67	11.75	.83	22.75	15.77
II	8.00	8.40	17.75	19.31	9.25	7.76	25.75	24.25
III	25.75	9.01	23.50	21.80	17.75	19.61	41.75	14.29
IV	50.00	—	50.00	—	50.00	—	50.00	—

The data were also analyzed using the original raw scores with the maximum of 72. The results were that the main effects for treatment and SES were significant at the .05 level and no other effects were significant. These results precisely replicate the results of the previous analysis. The modified data, however, were used in the tables and figure since they reduce the unnecessarily high scores of the nonlearners.

The performance of the subjects on the concept-formation task is displayed graphically in Figure 1. The high-SES subjects were noticeably superior to the low-SES subjects on all treatment conditions except Treatment Condition IV in which none of the subjects in either group reached criterion. That the difference between the SES groups is significant is indicated by the analysis of variance.

The effectiveness of the individual treatment conditions was analyzed with the Neuman-Keuls procedure for post hoc comparisons. Treatment Conditions I, II, and III are all superior to Treatment IV which contained essentially no training. In Treatment IV only the vowel–consonant discrimination task was administered prior to the concept-formation task. Since no subject in Treatment IV reached criterion, the effect of the vowel–consonant discrimination presented alone was negligible. It is noteworthy that Treatment III, which included the rule verbalization task as well as the vowel–consonant discrimination task, was superior to Treatment IV, which contained only the vowel–consonant discrimination task. Thus, learning to verbalize the rule markedly facilitated learning on the concept-formation task, provided that the vowel–consonant discrimination task was mastered. Since no test was made of the effect of the rule verbalization task in the absence of the vowel–consonant discrimination task, it is impossible to determine whether the effect of the rule verbalization is a main effect or an effect of the interaction between the rule verbalization and the presence of the vowel–consonant discrimination task. However, the rule contains the words "vowel" and "consonant"; a task analysis of the concept-formation task indicates that the discrimination learned in the vowel–consonant discrimination task is required for mastery on the concept-formation task. Consequently, it is highly likely that the rule verbalization would have little effect of the absence of mastery of the vowel–consonant discrimination. However, an empirical study is needed to verify this interpretation of the effect of rule verbalization.

Treatment Condition II was found to be superior to Treatment III. Since Treatment II included the rule application task in addition to the rule verbalization task, whereas Treatment III contained only the rule verbalization task, this effect is attributable to the application task. The question of whether this effect could have occurred without the rule application task having been preceded by the rule verbalization task does not arise since the act of verbalizing the rule is an indispensable step in the rule application procedure. The application of the rule cannot exist without the prior verbalization of the rule.

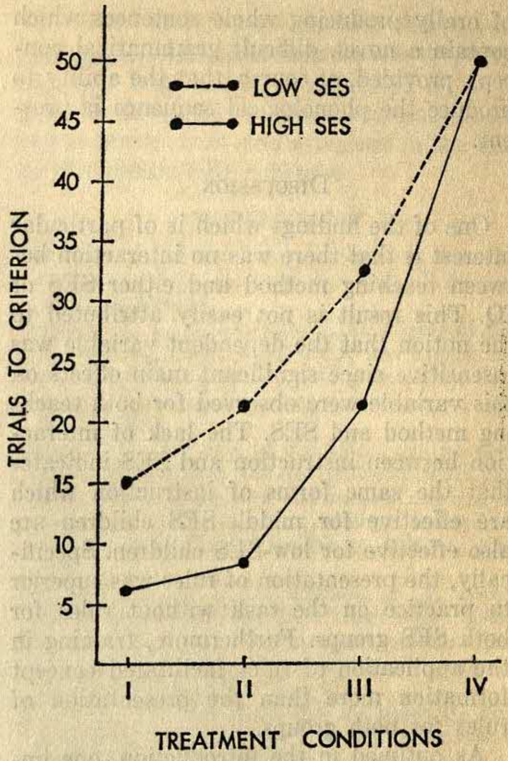


FIG. 1. Trials to criterion on the concept-formation task (Task 5) as a function of treatment conditions and SES.

Treatment Condition II, which included the rule verbalization task and rule application task which was appropriate for the concept-formation task, was not significantly different from Treatment Condition I. Treatment I contained the task which required subjects to produce instances of the concept in isolation (Task 4) and this task was preceded by the rule verbalization task and the rule application task appropriate for the production of instances. Since the rule application task in Treatments I and II differ slightly, these treatments do not bear the same relationship as Treatments II and III and Treatments III and IV. In the latter two pairs of treatments, the treatment with the lower Roman numeral contains a number of tasks which include one task not found in the treatment with the higher Roman numeral. Treatment I differs from Treatment II, however, not only in the addition of the task requiring the production of instances but also a modification of the rule application task. Thus, Treatments I and II contain two distinct differences rather than one. The lack of difference between these treatments is not easily attributable to a ceiling effect in the criterion measure since there is no sizable correlation between the means and variances of these groups. The two instructional sequences should be regarded as equally efficient methods for imparting the capability of orally producing whole sentences which contain a novel, difficult grammatical concept, provided, of course, that the ability to produce the phonological sequence is present.

DISCUSSION

One of the findings which is of particular interest is that there was no interaction between teaching method and either SES or IQ. This result is not easily attributed to the notion that the dependent variable was insensitive since significant main effects on this variable were observed for both teaching method and SES. The lack of interaction between instruction and SES indicates that the same forms of instruction which are effective for middle-SES children are also effective for low-SES children. Specifically, the presentation of rules was superior to practice on the task without rules for both SES groups. Furthermore, training in the application of rules facilitated concept formation more than the presentation of rules for both groups.

As outlined in the introduction, one implication of Jensen's research is that since low-SES children have substantial basic associative learning ability, instruction for this population should be developed which utilizes this ability and avoids unnecessary reliance on abstract verbal skills. In the present study, Treatment IV provided only practice on the criterion task with feedback. Under this treatment condition, learning depended heavily on basic associative learning ability and very little on abstract verbal functioning. The results indicated that under this condition little or no learning occurred for any subjects, and the two SES groups were not different from each other. Thus the low-SES group did not profit from the instructional strategy which depended on basic associative ability.

On the other hand, in Treatment III the subjects learned to verbalize an abstract rule before undertaking the concept-formation task. In this treatment condition, concept formation would be facilitated if the subjects could use the abstract verbal rule to respond appropriately to instances of the concept. Most of the subjects in this treatment condition attained the concept faster than subjects in Treatment IV. Consequently, it appears that concept formation is facilitated more by a method which requires the subjects to learn an abstract rule than by a method which requires basic associative ability regardless of the SES of the subjects.

The effect of intelligence on concept formation in this study was negligible. The main effect of IQ on concept attainment has been observed by some investigators (Osler & Fivel, 1961) and not by others (Heal, 1966; Rapier, 1967). A review of the research on concept formation and concept shift learning (Wolff, 1967) indicates that the role of intelligence in this area is inconsistent and not well understood. It should be added that since there was no interaction between treatment and IQ, the same conclusions may be drawn for the relationship of treatment and IQ as for the relationship of treatment and SES.

Although the rule verbalization treatment was superior to the treatment containing no rule in this experiment, this result does not completely replicate the finding in a prior study employing the same variables (Guthrie & Baldwin, 1970). Previously, the rule treatment produced no facilitation on the concept-formation task. One plausible explanation for these apparently conflicting results is that the difficulty of the concept-formation task in the two studies was different. In the previous experiment (Guthrie & Baldwin, 1970) the criterion task contained a maximum of 78 trials. The average number of trials required to reach criterion for the subjects who received no training on this task was 38.25. However, when no training was provided on the criterion concept-formation task in the present study, none of the subjects demonstrated any evidence of learning. Consequently, it appears that on the easier task, that is, when the task could be easily learned with no instruction, the provision of a rule did not affect performance. However, when the task was extremely difficult, the provision of the rule significantly improved learning.

The treatment condition which included both rule application and rule verbalization training was superior to the treatment which included only rule verbalization training. This result replicates the same finding in the previous study (Guthrie & Baldwin, 1970). The fact that the rule application training produced unambiguous facilitative effects in two experiments and the fact that these effects were observed for subjects regardless of their IQs or SES indicates the breadth and power of this variable.

It is possible that the SES effect observed in this study is due to the difference between the two SES groups in motivation, that is, effort, attention, and perseverence, rather than cognitive capabilities. Evidence for this interpretation is furnished by a study by Marshall (1969) which indicates that middle- and low-SES groups did not differ in learning an exciting, novel task which was interesting to both groups. On a low-interest task, however, the low-SES group performed worse than they had on the high-interest task, and the high-SES group performed better than they had on the high-interest task. In other words, a highly interesting task facilitated performance for the low-SES subjects and impaired the performance of the middle-SES group. If the task in the present study had been of greater intrinsic interest to the subjects, the SES effect may have been reduced or eliminated.

Further evidence for the motivational interpretation of these results is derived from research which indicates that different types of feedback and reinforcement are differentially effective for children from low- and middle-SES groups (Marshall, 1969; Zigler & DeLabry, 1962; Zigler & Kanzer, 1962). The use of praise and tangible reinforcers such as toys facilitates learning in low-SES children more than feedback consisting of information regarding the correctness of their responses. On the other hand, for middle-SES children informational feedback is more effective than tangible reinforcers or praise. It is likely that the low-SES subjects were not investing as much effort and attention in the task as the middle-SES subjects.

REFERENCES

CRONBACH, L. J. Heredity, environment, and educational policy. *Harvard Educational Review,* 1969, **39,** 338–347.

GUTHRIE, J. T., & BALDWIN, T. L. Effects of discrimination, grammatical rules, and application of rules on the acquisition of grammatical concepts. *Journal of Educational Psychology,* 1970, **61,** 358–364.

HEAL, L. W. The role of cue value, cue novelty, and overtraining in the discrimination shift performance of retardates and normal children of comparable discrimination ability. *Journal of Experimental Child Psychology,* 1966, **4,** 126–142.

HOLLINGSHEAD, A. B., & REDLICH, F. E. *Social class and mental illness.* New York: Wiley, 1958.

JENSEN, A. R. Social class, race and genetics: Implications for education. *American Educational Research Journal,* 1968, **5,** 1–42.

JENSEN, A. R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review,* 1969, **39,** 1–123.

MARSHALL, H. H. Learning as a function of task interest, reinforcement, and social class variables. *Journal of Educational Psychology,* 1969, **60,** 133–137.

OSLER, S. F., & FIVEL, M. W. Concept attainment: I. The role of age and intelligence in concept attainment by induction. *Journal of Experimental Psychology,* 1961, **62,** 1–8.

RAPIER, J. L. Effects of verbal mediation upon the learning of Mexican-American Children. *California Journal of Educational Research,* 1967, **18,** 40–48.

WOLFF, J. L. Concept-shift and discrimination-reversal learning in humans. *Psychological Bulletin,* 1967, **68,** 369–408.

ZIGLER, E., & DELABRY, J. Concept switching in middle class, lower class, and retarded children. *Journal of Abnormal and Social Psychology,* 1962, **65,** 267–273.

ZIGLER, E., & KANZER, P. The effectiveness of two classes of verbal reinforcers on the performance of middle and lower class children. *Journal of Personality,* 1962, **30,** 157–163.

(Received May 5, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 4, 352–356

EFFECT OF IRRELEVANT ATTRIBUTES AND IRRELEVANT OPERATIONS ON RULE LEARNING[1]†

JOSEPH M. SCANDURA[1] AND DONALD J. VOORHIES[2]
University of Pennsylvania

The negative effect of irrelevant attributes on concept learning has been well documented. Rule learning, however, involves *both* attributes and operations. The purpose of this study was to investigate the effects of both kinds of irrelevancies in rule learning. A 3 × 3 factorial design was employed, with the number of irrelevant attributes (0, 1, 2) and the number of irrelevant operations (0, 1, 2) being the factors. Both factors delayed rule learning. The results were compatible with the notion that subjects select from among the available attributes and operations, testing the various combinations in quasi-systematic fashion, until a rule is found which satisfies all instances.

The inhibiting effects of irrelevant attributes in paired-associate learning (e.g. Scandura, 1965; Underwood, 1963) and in concept learning (e.g. Archer, Bourne, & Brown, 1955; Bulgarella & Archer, 1962) are well known. Corresponding data in the area of rule learning, however, are almost entirely lacking. The reasons for this are not hard to find. For one thing, there has been relatively little research on rule learning. For another, there has been little general agreement on exactly what a rule is, and even less on the relationship between rules, on the one hand, and concepts and associations, on the other (e.g., see Gagné, 1965; Haygood & Bourne, 1965; Scandura, 1965, 1968, 1970).

While there is still no general agreement, the formulations of Gagné and Scandura appear to have much in common. For example, Gagné (1966) has viewed rules as a pair of concepts connected by an "action concept," and Scandura (1970) has characterized rules in terms of a set of stimulus attributes (determining properties) and an operation which acts on these attributes to generate a set of response attributes.

Scandura (1966, 1967, 1969) has argued that when rules are viewed in this way, concepts become simply special cases of rules in which the operation is degenerate. For example, the concept "red square" can be viewed in terms of the rule, "Say 'yes' when shown a stimulus which is a red square, and 'no' to all the others." In the case of associations, the operation is further restricted to a simple linking between one stimulus and its response.

In the case of concept learning, then, knowing the relevant attributes is tantamount to knowing the concept. Since the operation is of such a simple form, the question of which "operations" are to be used rarely arises. In rule learning, however, it is important to know both the relevant attributes and also the relevant operations. One might reasonably expect, therefore, that increasing the number of irrelevant operations would make it more difficult for a subject to determine which are relevant.

Indirect support for this contention was found in an experiment reported by Scandura (1966, 1969). In that study, the number of relevant and irrelevant attributes and operations remained constant, but cues indicating which attributes and/or operations were to be used were varied. Both types of cue had a facilitating effect on rule learning. This suggests that identifying the relevant attributes and operations in rule learning is not a trivial task where there are irrelevant attributes and operations from which to select.

---

[1] Requests for reprints should be sent to Joseph M. Scandura, Graduate School of Education, University of Pennsylvania, Philadelphia, Pennsylvania, 19104.

[2] The second author's participation was made possible by National Defense Education Act (IV) and United States Office of Education and research training fellowships in mathematics education research. The authors wish to thank Dorothy Jones, Louis Ackler, Francine Endicott, Jerry Satlow, and Chris Toy for their help with various facets of the study; Martin Hubley, Chairman of the Mathematics Department, Abington High School, South Campus, Abington, Pennsylvania, for his generous assistance in providing facilities and subects.

The primary purpose of this experiment was to test the dual hypothesis that both irrelevant attributes and irrelevant operations would delay rule learning. A secondary purpose was to obtain information on how subjects select attributes and operations on meaningful tasks. In particular, most theorists (Levine, 1966; Restle, 1962) have assumed that subjects select attributes on a random basis. This assumption seems reasonable in the experiments cited since the tasks were artificial and constructed explicitly so that each attribute appeared to the subject to be an equally likely candidate. It is not clear, however, that this assumption is realistic with meaningful materials with which the subjects have varying degrees of familiarity. In the Scandura (1966, 1969) study cited above, for example, the attributes and operations selected on learning trials appeared to be a function of the particular instances involved.

METHOD

Materials

The materials were similar to those used in an experiment reported by Scandura (1966, 1969). The stimuli were strings of three, four, and five single-digit natural numbers $(1, 2, 3, 4, \cdots, 8, 9)$. These stimuli were called 3-, 4-, and 5-tuples, respectively. The 3-tuples were of the form (A, B, C); 4-tuples, of the form (A, B, C, D); and 5-tuples, of the form (A, B, C, D, E), where in each case A, B, C, D, and E represent arbitrary natural numbers. Sample 3-, 4-, and 5-tuples, respectively, are (3, 2, 7), (5, 1, 4, 8), and (9, 4, 5, 1, 3).

A single natural number response was associated with each stimulus *n*-tuple according to a rule. Each rule involved the numbers in three positions of the associated stimulus *n*-tuples and two of the four arithmetic operations $(+, -, \times, \div)$. In accordance with Scandura's (1970) characterization, then, each rule can be characterized in terms of a set of triples of numbers (the set of determining properties—attributes) and a composite of two arithmetic operations. For example, the set of determining properties might consist of those triples of numbers (of, say, stimulus 5-tuples) appearing in the first, third, and fourth positions. The composite operation, in turn, might involve addition (applied, say, to the numbers in the first and fourth positions) followed by multiplication (applied to the obtained sum and the number in the third position). For example, application of this rule to the stimulus 5-tuple (9, 4, 5, 1, 3) yields $(9 + 1)5 = 50$ as the response.

A problem consisted of a sequence of 11 stimulus-response pairs all of a given type (e.g., all 4-tuples). The stimulus-response pairs were constructed (as indicated below) so that the responses could be generated from the associated stimuli by exactly one rule. Since each rule involved the numbers in just three positions, restricting the stimulus *n*-tuples to a given type had the effect of specifying a number of irrelevant attributes (IA_0, IA_1, IA_2). IA_0 refers to 3-tuples, where none of the numbers (positions) were irrelevant. IA_1 refers to 4-tuples, where the number in one position was irrelevant. IA_2 refers to 5-tuples, where the numbers in two positions were irrelevant. The number of arithmetic operations which could be used on a given problem was also restricted, this time by specifying a subset of the four operations from which the selections could be made. Where the selection set consisted of just two operations, there were no irrelevant operations (IO_0). Where the selection set consisted of three operations, one was irrelevant (IO_1); and where all four arithmetic operations were in the selection set, there were two irrelevant operations (IO_2).

Six rules were used as a basis for constructing the experimental problems (see Table 1). These rules were randomly selected subject to the following constraints (which helped to insure a broad sample of problems). (*a*) If $(X * Y) \# Z$ was used, $(X \# Y) * Z$ was not used (where * and # denote arbitrary operations). (*b*) $(X + Y) - Z$ and $(X \times Y) \div Z$ were included. (*c*) $(X + Y) \times Z$ and $(X - Y) \times Z$ were not both used. (*d*) $(X + Y) \div Z$ and $(X - Y) \div Z$ were not both used. (*e*) The operations on the left were always performed first (e.g., $(X * Y) \# Z$ was used but not $X * (Y \# Z)$). (X, Y, and Z refer to arbitrary positions in *n*-tuples. In one case, for example, X might refer to the second position, Y to the third position, and Z to the first.)

The stimuli for the 5-tuple problems were constructed first. For each of the six operational rules, 11 stimuli were constructed so that the natural numbers (between 1 and 9) assigned to the relevant positions in each *n*-tuple gave natural num-

TABLE 1

RULES USED TO CONSTRUCT EXPERIMENTAL PROBLEMS

3-tuple rules	4-tuple rules	5-tuple rules
$[(A \times B) - C]$	$[(A \times B) - C]$	$[(A \times B) - C]$
$[(B + C) - A]$	$[(B + D) - A]$	$[(B + E) - A]$
$[(B + C) \times A]$	$[(B + D) \times A]$	$[(B + D) \times A]$
$[(A \times C) \div B]$	$[(A \times C) \div B]$	$[(A \times C) \div B]$
$[(B \div C) + A]$	$[(B \div D) + A]$	$[(B \div D) + A]$
$[(C - A) \div B]$	$[(D - A) \div \mathrm{B}]$	$[(D - A) \div B]$

ber responses. The relevant positions and the order of operation on these positions were determined randomly, as were the numbers in the irrelevant positions. The stimuli for the 4- and 3-tuple problems were obtained by suppressing the irrelevant number(s) furthest to the right in the corresponding 5-tuples.

Each n-tuple was printed on the front of a 3 × 5 card; the same n-tuple, with the corresponding number response, appeared on the back. The selection set of operations associated with each problem was printed on a separate 3 × 5 card.

Design and Subjects

A 3 × 3 factorial design was used, the factors being the number of irrelevant attributes (IA_0, IA_1, IA_2) and the number of irrelevant operations (IO_0, IO_1, IO_2).

The subjects were 45 apparently motivated volunteer junior and senior students from Abington High School, South Campus, Abington, Pennsylvania. They were randomly assigned to the nine cells, so that five subjects were assigned to each cell.

Each subject received three problems. To help insure comparability of the nine experimental cells, the same pattern of assigning problems to subjects was used in each cell. This pattern of problems was randomly determined subject to two constraints: (*a*) No subject received the same problem twice. (There were only six problems from which to select.) (b) No problem appeared more than three times. The actual assignment of problems used is given in Table 2. (The n-tuples for the selected problems were always given in the same sequence; the starting point of that sequence was varied when the same problem was given more than once in any one cell. Across cells, the same sequence and starting points were used.)

Procedure

The subjects were tested individually during 45-minute study hall periods. Each subject was shown a sample n-tuple and selection set of operations, and was instructed to discover a rule (of the type described above) which would make it possible to correctly anticipate the responses. Concurrently, the subjects were provided with training in how to describe rules in words. The subject was told that there might be more than one way to generate the response corresponding to any given n-tuple, but that only one rule would generate the responses for all 11 n-tuples in a problem. (In fact, the chances were practically nil that more than one rule would work with as many as three successive stimulus-response pairs. In 100% of the problems, there was exactly one rule which could be used to generate the correct responses for the first three stimulus-response pairs. Further, only one rule could generate the correct responses for the first two stimulus-response pairs in 92% of the problems.)

TABLE 2

ASSIGNMENT OF RULES USED IN ALL NINE EXPERIMENTAL CELLS

Subject	Problem 1	Problem 2	Problem 3
1	$[(X \div Y) + Z]$	$[(X + Y) - Z]$	$[(X \times Y) - Z]$
2	$[(X + Y) \times Z]$	$[(X + Y) - Z]$	$[(X \div Y) + Z]$
3	$[(X \div Y) + Z]$	$[(X \times Y) \div Z]$	$[(X + Y) - Z]$
4	$[(X \times Y) \div Z]$	$[(X \times Y) - Z]$	$[(X - Y) \div Z]$
5	$[(X \times Y) - Z]$	$[(X + Y) \times Z]$	$[(X - Y) \div Z]$

To familiarize the subject with the procedure, he was presented with a practice problem prior to the three experimental problems. The following procedure was used throughout. The subject was shown the first n-tuple (stimulus) response pair, together with the second stimulus n-tuple, and was asked to give a response to the second n-tuple. As soon as the subject responded, the second card was turned over revealing the same n-tuple with its associated response. If the subject's response was incorrect, the examiner presented a new stimulus n-tuple, and asked for a response. If the subject's response was correct, the examiner asked what rule the subject was using. If the subject identified the correct rule, the examiner recorded the total time the subject spent on the problem. In computing the total time, the examiner ignored the time it took the subject to verbalize the rule after he gave correct responses. If not, the examiner told the subject that his rule was incorrect, and presented him with another stimulus n-tuple. This process was continued until the subject discovered the correct rule or 8 minutes had elapsed, whichever occurred first.

The subject was given the option of turning over a card to see the correct response at any time he wished. When this occurred, the examiner gave the subject a new stimulus n-tuple. If the subject spent 2 minutes with a stimulus n-tuple, the examiner urged him to move on to a new n-tuple. All previous stimulus-response pairs were visible to the subject throughout the problem.

The subject's score was the average time on the three problems he worked. After testing was completed, some of the subjects were presented with one or two additional problems and were asked to explain how they attacked the problems.

RESULTS AND DISCUSSION

As hypothesized, both irrelevant attributes and irrelevant operations delayed rule learning ($p < .001$). The group means are presented in Table 3. In spite of the small number of subjects (five) in each cell, the pattern of results was quite regular with the exception of cell IO_0–IA_1. This suggests that the variables considered were basically quite powerful.

Presumably, the subjects identified the appropriate rule on each problem by selecting and testing various combinations of attributes and operations until the correct rule was found. If the selection process was random, as many theorists have assumed (e.g., Levine, 1966; Restle, 1962), then the probability of selecting a correct rule can be determined for each cell of the design.

Assuming that the time required to sample (with replacement) and test each combination of attributes and operations is constant, the data of Table 3 should vary inversely with these probabilities. Equivalently, these data should vary directly with the numbers reported in Table 4 which were obtained by computing the number of instances required to attain a probability of .5 (an arbitrary figure) of selecting a correct rule and multiplying by a constant.

Comparison of Tables 3 and 4 indicates that while the general trends are in the same direction, there was an important difference. In Table 3, the difference between columns (rows) IA_1 (IO_1) and IA_0 (IO_0) was substantially greater than the corresponding difference between columns (rows) IA_2 (IO_2) and IA_1 (IO_1). In Table 4, precisely the opposite is true. In effect, the subjects did much better on the IA_2 and IO_2 problems than would be expected on the basis of random selection.

This observation suggests that to some extent, at least, the subjects tended to make their selections in a systematic fashion. Indirect evidence for this contention was obtained from the postexperimental questioning. Most of the subjects indicated that various characteristics of the stimuli and responses caused them to either eliminate or single out for special attention certain attributes and operations. For example, whenever a large response (number) was associated with a stimulus, many of the subjects tended to try multiplication first, and possibly addition second. Similarly, small responses resulted in the use of division, or sometimes subtraction.

TABLE 3

MEAN TIME (IN SECONDS) TO PROBLEM COMPLETION IN THE NINE EXPERIMENTAL CELLS

Irrelevant operations	IA_0	IA_1	IA_2
IO_0	65	155	101
IO_1	91	180	200
IO_2	110	226	293

TABLE 4

PREDICTED PATTERN OF RESULTS

Irrelevant operations	IA_0	IA_1	IA_2
IO_0	1	4	10
IO_1	3	12	30
IO_2	6	24	60

Note.—The numbers in this table are meaningful onlv in relation to one another.

It must not be thought from the results of this experiment that the effects of irrelevant attributes and irrelevant operations are uniformly negative. While not dealt with in this study, it is quite possible that the introduction of such irrelevant information may cause subjects to search for new and better ways of discovering rules. To the extent that they succeed (cf. Roughead & Scandura, 1968), irrelevant information may enhance transfer to new problems. This is an important possibility which should be checked in further research.

REFERENCES

ARCHER, E. J., BOURNE, L. E., & BROWN, F. G. Concept identification as a function of irrelevant information and instructions. *Journal of Experimental Psychology*, 1955, **49**, 153–164.

BULGARELLA, R. G., & ARCHER, E. J. Concept identification of auditory stimuli as a function of amount of relevant and irrelevant information. *Journal of Experimental Psychology*, 1962, **63**, 254–257.

GAGNÉ, R. M. *The conditions of learning*. New York: Holt, Rinehart & Winston, 1965.

GAGNÉ, R. M. The learning of principles. In H. J. Klausmeier & C. W. Harris (Eds.), *Analysis of concept learning*. New York: Academic Press, 1966.

HAYGOOD, R. C., & BOURNE, L. E. Attribute- and rule-learning aspects of conceptual behavior. *Psychological Review*, 1965, **72**, 175–195.

LEVINE, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, 1966, **71**, 331–338.

RESTLE, F. The selection of strategies in cue learning. *Psychological Review,* 1962, **69,** 329–343.

ROUGHEAD, W. G., & SCANDURA, J. M. "What is learned" in mathematical discovery. *Journal of Educational Psychology,* 1968, **59,** 283–289.

SCANDURA, J. M. Familiarization and cue selection. *Psychological Reports,* 1965, **17,** 27–38.

SCANDURA, J. M. Precision in research on mathematics learning: The emerging field of psychomathematics. *Journal of Research in Science Teaching,* 1966, **4,** 253–274.

SCANDURA, J. M. The basic unit in meaningful learning—Association or principle? *The School Review* 1967, **75,** 329–341.

SCANDURA, J. M. New directions for theory and research on rule learning: I. A set function language. *Acta Psychologica,* 1968, **28,** 301–321.

SCANDURA, J. M. New directions for theory and research on rule learning: II. Empirical research. *Acta Psychologica,* 1969, **29,** 101–133.

SCANDURA, J. M. The role of rules in behavior: Toward an operational definition of "what (rule) is learned." *Psychological Review,* 1970, **77,** 516–533.

UNDERWOOD, B. J. Stimulus selection in verbal learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw Hill, 1963.

(Received May 24, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 4, 357-364

INTERACTION AMONG ANXIETY, STRESS, RESPONSE MODE, AND FAMILIARITY OF SUBJECT MATTER ON ACHIEVEMENT FROM PROGRAMMED INSTRUCTION[1]†

SIGMUND TOBIAS AND THEODORE ABRAMSON
City College, City University of New York
Division of Teacher Education, City University of New York

This study intended to establish an Attribute × Treatment interaction among anxiety, stress, and response mode to programmed instruction. A total of 144 subjects were randomly assigned either to a constructed response mode, with or without reinforcement, or to a reading condition. Stress instructions were used for half the sample. Multiple linear regression analysis indicated that the reinforced constructed response group achieved more (than the reading group) on technical, difficult material. A significant interaction among facilitating anxiety and other variables was interpreted in terms of achievement motivation. Debilitating anxiety interacted with stress on easy materials, but did not interact with other variables on complex, technical content. Implications of these results for anxiety research in similar settings are discussed.

The adaptation of instructional methods to individual differences in aptitudes among pupils is an area of some importance to education. In order to assign a pupil to the instructional method which would result in optimal achievement for him, an interaction is presumed to exist between the instructional strategy and the aptitude. The study of this problem has become known as the Aptitude × Treatment interaction (ATI), or, as recently suggested (Tobias, 1969c), the Attribute × Treatment interaction.

In an extensive review of ATI research, Cronbach and Snow (1969) concluded that a substantial research effort was needed in order to establish ATIs as a viable construct in the area of instructional research. A review of ATI studies in which programmed materials were used (Tobias, 1969c) suggested that among a number of possible reasons for not finding ATIs two were most probable: (*a*) the failure to demonstrate that the variable measured by test score was actually operative in the experimental situation; (*b*) the failure to vary subject matter since ATIs may appear in one type of content and not another. In the present study, stress instructions were employed in order to make it more likely that the anxiety measured by tests was actually operative in the research situation, and the content characteristics of the material were varied by employing materials of differing familiarity.

The instructional treatment manipulated in this experiment was the mode of responding to the programmed materials. There have been discrepant findings regarding the effects of response modes on achievement (Anderson, 1967). Constructing responses and then checking their accuracy has not consistently led to higher achievement, compared to reading the program with the responses filled in, or thinking the answer without actually constructing it.

Previous research (Tobias, 1969a, 1969b) suggested that constructing responses does not lead to superior achievement on materials with which subjects have had prior experience, while achievement on technical, unfamiliar material was superior when subjects constructed their responses. Theoretically, it is not unreasonable to assume that content with which subjects are unfamiliar may require a more overt response for optimal learning than does familiar subject matter. Lack of familiarity implies that the responses required by the task may not be in the subject's repertory, and consequently, actually responding overtly may be the best way of mastering it. Whereas for familiar subject matter, the required responses may be in the subject's repertory—though perhaps not in the exact context required by the material—and thus, overt responding may not be as important. On such tasks the subject does not have to learn how to make the responses, but rather how to connect existing responses to new, or different, situations. Overtly making a response which is already in the repertory may not strengthen its association to new situations to any measurable degree more than covertly thinking or reading the material. This formulation suggests that for content in which the required responses have been previously learned little achievement difference is to be expected between overt or covert responding, choosing from one of several alternatives, or reading the material. When the responses required by a program are new, actually making an overt response and receiving knowledge of results concerning it, is likely to lead to superior achievement.

---

[1] This study was supported by a grant from the United States Office of Education, Bureau of Research Project No. 8-b-015. A report on this project was read at the American Educational Research Association Annual Convention, Minneapolis, March 1970.

[2] Requests for reprints should be sent to Sigmund Tobias, Klapper Hall 6, City College, City University of New York, New York 10031.

This familiarity interpretation is in accord with the results of Daniel and Murdoch (1968) who found overall achievement differences favoring a group studying operant psychology with a constructed response program (Holland & Skinner, 1961) compared to subjects studying the same subject with nonprogrammed writings. When differences between the groups on the most technical, and presumably least familiar, subtest were eliminated from the dependent measures by covariance adjustment the other group differences were no longer significant.

Similarly, Roderick and Anderson (1968) found that high school students who used the Holland and Skinner program achieved more than subjects studying the material with succinct textbook-like passages. However, for college subjects who were presumably more familiar with this topic, there were no achievement differences between the two formats. These studies support the hypothesis that the optimal mode of responding to programmed materials may depend on the familiarity subjects have with the content. Thus, a second purpose of this study was to replicate previous results concerning the familiarity formulation.

Anxiety

There is a strong rationale for the relationship between anxiety and achievement from programmed instruction. Specifically, it would seem that the tight organization, reduction of uncertainty, and high ratio of reinforcement which characterize most programmed materials should be especially advantageous for the achievement of anxious students. It is, therefore, surprising to note that results relating anxiety to programmed instruction, reviewed elsewhere (Tobias, 1969c; Tobias & Williamson, 1968), largely fail to support these expectations. Campeau (1968) did find an interaction between anxiety and feedback in programmed instruction for girls but not for boys. High-anxiety girls achieved more than a low-anxiety group in a constructed response with reinforcement condition. When reinforcement was removed the achievement of the lower anxiety subjects was higher. No data concerning performance on the program, such as error rates or time required to complete the program, were reported. In the absence of such information inferences regarding the degree to which anxiety was operative during the research situation are speculative.

Investigations relating anxiety to programmed instruction have typically sought relatively simple effects. Generally, an anxiety scale was administered, and different programming formats implemented. Prior investigations have not taken account of some of the complexities reported in experimental investigation of the relationship between anxiety and learning—such as the relationship between anxiety and subject-matter difficulty. According to the drive theory proposed by Spence (1958) and Taylor (1956), high anxiety should interfere with performance on complex tasks for which the subject's incorrect responses are of equal, or higher, strength than the correct responses. For simpler tasks, where one predominant response exists relative to other responses, anxiety is predicted to have a facilitating effect.

Another frequently reported finding in the anxiety literature is an interaction between anxiety and stress. Differences between high- and low-anxiety groups occurred only under conditions of stress (Nicholson, 1958; Sarason, 1958; Sarason & Palola, 1960). When stress was absent, no anxiety effects were observed. As previously suggested, these findings reemphasize the importance of coupling a variable assigned on the basis of test score with an experimental variation to increase the probability that the variable studied is operative in the research situation.

A further purpose of this study was to investigate some of these anxiety effects in a meaningful learning context using programmed materials. Specifically, it was expected that there be no main effects attributable to anxiety but, instead, that anxiety would interact with situational stress. Furthermore, a triple interaction was predicted between anxiety, stress, and response mode. Constructed response was expected to be least affected by anxiety and stress, and providing no reinforcement most affected by it. Finally, the anxiety effects were expected to decrease learning on difficult content and increase achievement on simple material.

METHOD

The basic model for this investigation was similar to those recommended by Cronbach and Snow (1969) for ATI studies. Two independent variables were manipulated: stress and response mode. The degree to which these variables interacted with sex and with anxiety assigned on the basis of test scores was determined by multiple linear regression techniques.

Subjects

A total of 144 subjects, 80 of whom were female, participated in this study. Subjects were recruited primarily from educational psychology classes at the City College of New York during the fall 1967 and spring 1968 semesters. Subjects were told that the purpose of the experiment was to study the relationship between programmed instruction and the way people think, and were paid $6 for their participation.

Materials

The instructional materials were employed in prior investigations (Tobias, 1969a, 1969b) and contained both familiar and technical sections. The familiar portion of the program contained 54 frames dealing with the incidence of heart disease, and risk factors for contracting heart disease such as high cholesterol, smoking, age, etc. The technical, unfamiliar section of the program, 89 frames, dealt mainly with the diagnosis of myocardial infarction from the fifth precordial lead of the electrocardiogram. Technical names for different degrees of severity of coronary disease, electrocardiographic tracings characteristic of each level of severity, and graphic representations of the damage to the heart muscle caused by the various levels of coronary disease were included in this part of the program. Program difficulty, determined by the percentage of correct responses to the program, was 96.6% and 81.3% for the familiar and technical sections, respectively, a difference significant beyond the .001 level (Tobias, 1968). The instructional materials were presented in a booklet format with three frames to a page. The confirmation for one frame appeared in the left-hand margin of the succeeding frame on the next page for the constructed response group, and was eliminated for the no reinforcement groups.

A posttest had been previously developed covering the familiar and technical material. The familiar subtest had an alpha reliability of .66. Alpha reliabilities of two technical subtests were (*a*) .86 for the items dealing with content requiring the subject to respond with drawings, and (*b*) .85 for the verbal items. For the present investigation the latter two subscales were combined. The complete constructed response program, posttest, and criteria for scoring those responses requiring drawings of electrocardiogram tracings can be found in a previous report (Tobias, 1968).

The anxiety measure utilized in this investigation was Alpert and Haber's (1960) Achievement Anxiety Test (AAT). The AAT is composed of two subscales: the AAT+ deals with the kind of anxiety that facilitates performance in achievement situations, and the AAT− items tap the debilitating effects of anxiety on performance in achievement situations.

Procedures

Administration of the procedures of this study required two sessions. In the first session the anxiety measures, a pretest for the familiar programmed material, and some other research scales were administered. The instructional program and posttests were administered in the second session.

Half the subjects who volunteered for this study, and were classified as high or low anxious by a medial split of the AAT−, were randomly assigned to a stress condition. Stress was induced by instructions relating achievement on the program to intelligence, in general, and to learning ability in college, specifically. Within the stress condition, subjects were randomly assigned to one of three presentation modes: two constructed response groups, one with and one without reinforcement, and a reading group. Of the 72 subjects in the stress condition 70 were tested at one time early in the fall 1968 semester, in order to minimize the possibility of feedback between the stress and nonstress groups. Nonstress subjects were typically tested in small groups ranging from 1 to 10, with the single largest administration consisting of 22 subjects at one time.

RESULTS

The critical dependent measures in this investigation were the scores attained on the posttest. In order to allow for direct comparability between the means of the technical and familiar subtests, which contained a different number of raw score points, scores were converted to percentages.

The data were analyzed using the multiple linear regression techniques outlined by Cronbach and Snow (1969) and by Overall and Spiegel (1969). A subject's group membership in the experimental stress and response mode conditions was represented by a series of binary vectors. The three response mode conditions were expressed by two vectors: constructed response was coded 1 and 0; no reinforcement, 0 and 1; and reading, −1 for both vectors. This coding made it possible to replicate previous findings by comparing the constructed response and reading groups with the use of the first response mode vector (Cohen, 1968). The AAT scales were coded as continuous vectors. Interaction vectors between the experimental variables, and among these and the assigned variables were the simple products of the component vectors. Since previous research had frequently reported an interaction between sex and anxiety a binary sex vector was added.

TABLE 1

SUMMARY OF MULTIPLE LINEAR REGRESSION RESULTS OF ACHIEVEMENT DATA

Effect	*df*	Familiar		Technical	
		% variance	*F*	% variance	*F*
Response mode (A)	2			.03	2.88*
Constructed response vs. reading	1			.03	5.21**
Stress (B)	1	.01	1.37		
AAT− (C)	1				
AAT+ (D)	1	.01	1.14	.03	5.43**
Sex (E)	1	.01	1.20		
D × E	1			.08	8.41***
A × C	2				
A × B	2			.01	1.02
A × D × E	2				
B × C	1	.03	4.01*		
B × D × E	1				
A × B × C	2	.04	2.65	.01	1.05
A × B × D × E	2			.03	3.14**

Note.—*F* values less than 1 not shown.
* $p = .058$, probabilities determined by exact procedure.
** $p < .05$.
*** $p < .01$.

The initial analysis sought to determine whether anxiety and stress affected the sexes differentially. Interaction vectors were thus developed between sex and the AAT vectors, and between sex and stress. This analysis indicated that there were differential sex effects only for the AAT+ ($F = 12.90$, $p < .001$). Therefore, succeeding analyses of all AAT+ effects were conducted within sex by including the sex, and AAT+ × Sex vectors. The full model[3] for the analysis of achievement data included the vectors mentioned above, plus the main effects and only those interaction vectors which were of interest in this investigation. The vectors of interest and the order in which they were tested are shown in Table 1.

The succeeding analysis followed a modified stepdown procedure similar to the one described by Overall and Spiegel (1969), and by Cohen (1968). The significance of main effects was tested by forming a reduced model containing all the main effects, then testing for the significance of each variable by dropping that vector from the model, and testing for reduction in the multiple correlation. This procedure was followed since some of the main effects were intercorrelated (e.g., r between AAT+ and AAT− was .36) and allowed for the estimation of the percentage of variance contributed uniquely by any variable adjusted for the effects of all other variables. The second modification of the stepdown procedure was to employ the full model, containing all of the vectors including those testing for sex interaction, subjects' Scholastic Aptitude Test and pretest scores for a total of 23 *df*, in the denominator, rather than only the previous restricted model. This procedure resulted in a more conservative test than is recommended by Cronbach and Snow (1969). The interaction effects were examined by adding vectors in the order in which they appear in Table 1, and their significance tested by comparing them to the prior model, and dividing with the full model in the denominator.

---

[3] Beta weights, regression coefficients, and other data pertaining to the full models have been deposited with the National Auxiliary Publications Service. Order Document Number 00237 from National Auxiliary Publications Service, the American Society for Information Science, % CCM Information Sciences, Inc., 22 West 34th Street, New York, New York 10001. Remit in advance $5.00 for photocopies or $2.00 for microfiche.

As expected, Table 1 indicates that none of the main effects achieved significance. The interaction between the AAT− and stress was significant at the 4% level. Inspection of the beta weights for this effect indicates that in this condition anxiety interacted with stress to raise achievement.

For technical subject matter the response mode main effect is of borderline significance, its exact probability being .058. For the comparison between constructed response and reading groups, $F = 5.21$, $p = .02$. Inspection of Table 2 indicates that the constructed response mode resulted in significantly higher achievement on technical material than that attained by the reading group. This analysis, furthermore, suggests that the borderline significance for the response mode main effect is attributable largely to the presence of the no reinforcement group whose performance fell between that of the other groups. Contrary to expectation, none of the results involving debilitating anxiety, as measured by the AAT−, achieved significance. The AAT+ did have both a significant main effect, and interaction with response mode and stress for the technical material. The analysis reported for familiar and technical material was also computed for the amount of time taken to complete the program. This analysis yielded a huge main effect, $F = 155.7$, for response mode. Inspection of the means in Table 2 indicates that these results are attributable to the large difference between the two constructed response modes and the reading group with respect to the amount of time taken. None of the other effects were significant for the time results.

TABLE 2

MEANS AND STANDARD DEVIATIONS, IN PERCENTAGES, FOR FAMILIAR AND TECHNICAL SUBJECT MATTER FOR DIFFERENT RESPONSE MODES

Dependent measure	Constructed response		No reinforcement		Reading	
	M	*SD*	*M*	*SD*	*M*	*SD*
Familiar test	63.7	13.7	63.8	12.0	63.8	11.6
Technical test	63.1	14.3	59.9	13.9	57.4	16.5
Time on program	86.8	19.1	86.9	17.1	35.2	12.0
Familiar program	96.5	3.6	96.1	3.2	—	
Technical program	78.8	9.8	71.6	12.0	—	

Data for the percentage of correct responses to the program were available only for the two constructed response groups. These data were analyzed in the same manner as the achievement results. This analysis for the technical portion of the program yielded only one significant effect, namely for the response mode variable ($F = 13.24$, $p < .001$). The mean percentages for the two response groups, reproduced in the last row of Table 2, indicate that the constructed response group with reinforcement had a significantly lower error rate on the technical material than did the no reinforcement group. For the familiar material there was no significant difference between the groups. Neither the anxiety scales nor the stress condition, singly or in interaction, appeared to have a significant effect on the percentage correct on the program.

DISCUSSION

The results of this experiment largely fail to support the expected ATI between achievement from programmed instruction, stress, and debilitating anxiety, though an interaction with facilitating anxiety was found. Previous findings dealing with the effects of response mode to programmed instruction were replicated. The implications of these data are discussed below.

The findings that constructing responses leads to higher achievement only on technical unfamiliar content strengthens the response learning interpretation. The significance of the findings is obscured by the achievement data for the no reinforcement group which fell between the two other groups. It had been assumed that due to its difficulty this condition would be especially advantageous in a study involving anxiety and stress. This expectation was not supported. Interpretation of the data for the no reinforcement condition in terms of response learning is somewhat complex. An overt response was required, though the correct answer was not provided. Despite the fact that explicit reinforcement had been removed, it seems clear that implicit confirmation was present. The hierarchal organization, and frequent review of the content of prior frames present in linear programs suggests that by going from frame to frame subjects probably receive implicit reinforcement for their responses.

Time

In accord with the findings of other investigations present results indicate that both the constructed response groups took significantly longer than the reading group to cover the same material. In this study overt responding took 2½ times longer than reading the material. It is sometimes suggested (Roderick & Anderson, 1968) that such findings raise serious questions regarding the efficiency of programmed materials. A comparison of time required to achieve mastery is meaningful in those instances when the time required to learn a subject is of equal, or greater, importance than the mastery attained. On the other hand, asking subjects to skim through materials they have read once may improve their mastery of the subject matter to the same level as working an instructional program, and conceivably at less time. A study of that kind, involving different types of content, might be very revealing with respect to the efficiency issue.

Anxiety, Stress, and Programmed Instruction

The expected ATI between debilitating anxiety, stress, and response mode to programmed instruction was unsupported for technical materials, though stress did interact with anxiety to improve achievement on easy, familiar content.

Facilitating anxiety. The AAT+ had a strong main effect, and interacted with stress and response mode. It is difficult to interpret this finding in terms of anxiety theory. Alpert and Haber (1960) conceptualized this scale as a measure of anxiety which facilitates people's performance in achievement situations. The scale contains such items as the following: "The more important the exam or test, the better I seem to do." "I enjoy taking a difficult exam more than an easy one." "Nervousness while taking a test helps me to do better." In view of the content of these items it is not surprising to note that McKeachie (1969) suggested that the AAT+ may be more closely related to achievement motivation than to traditional conceptions of anxiety. These data suggest that the AAT+ may be a fruitful measure for use in ATI investigations studying academic motivation. Examination of the AAT+ × Sex interaction indicates that this scale is best used with female subjects since AAT+ score correlated .50 with technical achievement for females, and −.03 for males.

Stress and debilitating anxiety. The results of this study failed to confirm ATIs between anxiety, stress, and response mode to programmed instruction for difficult materials. One possible explanation of these data is that despite the induction of stress, anxiety was not clearly operative in the situation. The findings of no differences between the stress groups on either the program or posttest support this possibility. The best way of ascertaining that anxiety is operative in the research task may be to measure it while subjects are working on the materials by interspersing a brief anxiety scale with the instructional materials (O'Neill, Speilberger, & Hansen, 1969). Attributing the findings to the possibility that anxiety was not sufficiently engaged cannot account for the presence of an interaction between the AAT− and stress for the familiar, easy materials.

A more likely explanation of the present data is that the technical material was not difficult enough to evoke, or maintain, anxiety. The error rate for the technical material was 24.8% compared to 3.7% for the familiar ($t = 14$). In the O'Neill et al. (1969) study, in which interactions between task difficulty and anxiety were found, the error rate for the two-part difficult materials was 73% and 60%, compared to a virtual zero error rate for the easy materials. Their difficult task was made more complex since subjects could not go to the next problem without solving the prior one. Such levels of difficulty are virtually impossible to attain with programmed materials, since, while useful for research, they are likely to be useless for teaching instructional materials.

The present findings, combined with the negative results of other attempts to relate achievement from programmed instruction to anxiety, raises doubt whether such an interaction is to be expected in the context of programmed, or similar types, of instruction. If the ratio of right to wrong responses must be about 2–3:1 in order to demonstrate ATIs with anxiety, few instructional methods would qualify for such an interaction. The evidence suggests that the absolute error rates may well have to be in the 60% range in order to find evidence of ATIs with anxiety. In turn, while such ATIs may be of considerable theoretical interest, they may not be pertinent to the design of optimal instructional strategies.

This study offers relatively little support for ATIs between anxiety and programmed instruction. Taken together with the negative findings reported in other ATI investigations (Cronbach & Snow, 1969) little evidence exists for the interaction between attributes measured on the basis of test scores and performance on learning tasks. This study does indicate, however, that the subject's prior familiarity with the content is a useful differentiating variable in the prediction of optimal instructional strategy. These findings suggest that familiarity may well be a promising variable in the investigation of instructional strategies other than response mode. It appears possible that the subjects' prior experience with a particular content may prove to be more useful in the prediction of optimal instructional strategies than the attribute measures employed in typical ATI investigations. The use of previous familiarity as a differentiating variable should, however, await further demonstrations of its usefulness in interaction with other instructional strategies.

REFERENCES

ALPERT, R., & HABER, R. Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology,* 1960, **61,** 207–215.

ANDERSON, R. C. Educational psychology. *Annual Review of Psychology,* 1967, **18,** 103–164.

CAMPEAU, P. C. Test anxiety and feedback in programmed instruction. *Journal of Educational Psychology,* 1968, **59,** 159–163.

COHEN, J. Multiple regression as a general data analytic system. *Psychological Bulletin,* 1968, **70,** 426–443.

CRONBACH, L. J., & SNOW, R. E. *Individual differences in learning ability as a function of instructional variables.* (Final Report, United States Office of Education Contract No. OEC 4-6-061269-1217). Stanford, Calif.: Stanford University, 1969.

DANIEL, W. J., & MURDOCH, P. Effectiveness of learning from a programmed text compared with a conventional text covering the same material. *Journal of Educational Psychology,* 1968, **59,** 425–431.

HOLLAND, J. G., & SKINNER, B. F. *The analysis of behavior.* New York: McGraw-Hill, 1961.

MCKEACHIE, W. J. Interaction of achievement cues and facilitating anxiety in the achievement of women. *Journal of Applied Psychology,* 1969, **53,** 147–148.

NICHOLSON, W. M. The influence of anxiety upon learning: Interference or drive increment. *Journal of Personality,* 1958, **26,** 303–319.

O'NEILL, H. F., JR., SPIELBERGER, C. D., & HANSEN, D. N. The effect of state-anxiety and task difficulty on computer-assisted learning. *Journal of Educational Psychology,* 1969, **60,** 343–350.

OVERALL, J. E., & SPIEGEL, D. K. Concerning least squares analysis of experimental data. *Psychological Bulletin,* 1969, **72,** 311–322.

RODERICK, M., & ANDERSON, R. C. Programmed introduction to psychology versus text-book style summary of the same lesson. *Journal of Educational Psychology,* 1968, **59,** 381–387.

SARASON, I. Effects on verbal learning of anxiety, reassurance, and meaningfulness of material. *Journal of Experimental Psychology,* 1958, **56,** 472–477.

SARASON, I. H., & PALOLA, E. H. The relationship of test and general anxiety, difficulty of task, and experimental instruction to performance. *Journal of Experimental Psychology,* 1960, **59,** 185–191.

SPENCE, K. W. A theory of emotionally based drive (D) and its relation to performance in simple learning situations. *American Psychologist,* 1958, **13,** 131–141.

TAYLOR, J. A. Drive theory and manifest anxiety. *Psychological Bulletin,* 1956, **53,** 303–320.

TOBIAS, S. *The effect of creativity, response mode,* *and subject matter familiarity on achievement from programed instruction.* New York: MSS Educational Publishing Co., 1968.

TOBIAS, S. Effect of creativity, response mode, and subject matter familiarity on achievement from programmed instruction. *Journal of Educational Psychology,* 1969, **60,** 453–460. (a)

TOBIAS, S. *Distraction and programmed instruction.* (Technical Report No. 2, Programmed Instruction Research Project). New York: City College, City University of New York, 1969. (b)

TOBIAS, S. Research strategy in the effect of individual differences on achievement from programed instruction. Paper presented at the annual convention of the American Psychological Association, Washington, D.C., September 1969. (c)

TOBIAS, S., & WILLIAMSON, J. H. Anxiety and response mode to programed instruction. Paper presented at the American Educational Research Association Convention, Chicago, February 1968.

(Received May 24, 1970)

JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 62 October 1971 Number 5

CONTENTS

Transfer Variables and Sequence Effects in Subject-Matter Learning........... 365
STANLEY L. DENO, JOSEPH R. JENKINS, AND JUDY MARSEY

Effect of Incentive Variables and Type of Adjunct Question upon Text Learning.. 371
LAWRENCE T. FRASE

Elaboration Training and Paired-Associate Learning Efficiency in Children.... 376
WILLIAM D. ROHWER, JR. AND MARY SUE AMMON

Teaching Disadvantaged Preschool Children to Think Creatively with Pictures.. 384
JOE KHATENA

Effects of Three Types of Inserted Questions on Learning From Prose........... 387
GRAEME H. WATTS AND RICHARD C. ANDERSON

Meaningful Processing of Sentences...................................... 395
RICHARD C. ANDERSON, SHEILA R. GOLDBERG, AND JANET L. HIDDE

Effects of Variations in Content and Stimulus Grouping on Visual Sequential Memory.. 400
JOHN R. BERGAN, BARRY J. ZIMMERMAN, AND MAUREEN FERG

Comparative Study of the Perceived Relevance of Material to be Learned and its Impact on the Performance of Culturally Deprived Junior College Students.. 405
LAWRENCE R. O'LEARY

Empirical Test of the "Teacher-Centered" versus "Student-Centered" Dichotomy.. 410
FRANK COSTIN

Effects of Memory Support on State Anxiety and Performance in Computer-Assisted Learning.. 413
BARBARA L. LEHERISSEY, HAROLD F. O'NEIL, JR., AND DUNCAN N. HANSEN

Effects of Training Black Preschool Children in Vocabulary versus Sentence Construction.. 421
PAUL R. AMMON AND MARY SUE AMMON

Interactive Effects of Achievement Orientation and Teaching Style on Academic Achievement.. 427
GEORGE DOMINO

Effects of Verbal Warning upon Misapplication of a Rule of Limited Applicability.. 432
DUNCAN MCQUARRIE AND ARDEN GROTELUESCHEN

Verbal Facilitation of Paired-Associate Learning: A Limited Generalization... 439
JOEL R. LEVIN, JAMES W. HORVITZ, AND SANDRA A. KAPLAN

Erratum... 444

© 1971 by the American Psychological Association, Inc.

Journal of Educational Psychology
1971, Vol. 62, No. 5, 365–370

TRANSFER VARIABLES AND SEQUENCE EFFECTS IN SUBJECT-MATTER LEARNING

STANLEY L. DENO[1]
University of Minnesota
JOSEPH R. JENKINS AND JUDY MARSEY
University of Delaware

The present experiment was designed to study (*a*) the transfer of effects of learning a strategy (content-nonspecific transfer) or prerequisite subject matter (content-specific transfer), and (*b*) the effects of inductive–deductive sequences on learning complex subject-matter concepts in isolation and in prose context. One hundred thirty-two college sophomores participated as subjects in 11 different treatment and control groups. Analysis of variance and *t* tests were used in the data analysis. The results failed to demonstrate any advantage for content-nonspecific prior training. However, both content-specific prior training and deductive sequencing produced significant transfer in concept identification and on learning from a prose passage.

Research on the facilitating effects of prior learning on subsequent performance can be divided into two separate and fairly distinct categories. The first of these categories is exemplified by the research of Gagné and his associates (Gagné, 1965) and can be described as research on content-specific transfer where facilitation is presumed to be the effect of prior learning of specific content which is prerequisite or subordinate to the content of the transfer task. The second category of research on transfer of learning is exemplified by the work of Bruner, Goodnow, and Austin (1956) and is described as research on content-nonspecific transfer where facilitation is presumed to be the effect of acquiring a general learning strategy or style.

Not surprisingly, the work on content-specific transfer has grown out of research on subject-matter learning where the elaborate structure of relationships among concepts and principles within the subject matter becomes an important variable in investigation of learning efficiency (Carroll, 1964; Suppes, 1966). Research on learning strategies and styles on the other hand, emanates more directly from laboratories of experimental psychology where specific subject learning is usually of less concern. Research on both content-specific and content-nonspecific transfer of learning are distinct from research on transfer which is presumed to be the effect of specific or general learning abilities (e.g., Jensen, 1965) in that the implication of the first two is that the basis for transfer is educable while in the case of research on learning abilities, the same implication is not clear.

Educational psychologists have attempted to organize the findings of research on tranfer of learning into a coherent set of recommendations for teachers who are attempting to arrange optimum conditions for learning in their classrooms (see, e.g., DeCecco, 1968). The resulting prescriptions usually include recommendations for utilizing both content-specific and content-nonspecific transfer. While such recommendations are logically derivable from existing research, the lack of empirical data on the learning of complex subject-matter concepts makes most of these recommendations speculative (Carroll, 1964).

The present study was designed to determine the effects of both content-specific and content-nonspecific transfer variables in initial concept learning and subsequent learning from prose. More precisely, the research was designed to answer the following questions:

1. Does practice at classifying or learning a strategy add to the transfer effects of learning prerequisite content when learning to identify examples of a veridical concept?

---

[1] Requests for reprints should be sent to Stanley L. Deno, Special Education Department, University of Minnesota, 109 Pattee Hall, Minneapolis, Minnesota 55455.

Copyright © 1971 by the American Psychological Association, Inc.

2. Does training on identifying and labeling the attributes of a veridical concept improve performance on identifying examples of that concept?

3. Does prerequisite concept learning improve subsequent prose learning involving those concepts?

In an effort to shed some light on the continuing debate over the relative effectiveness of inductive and deductive approaches to learning (Guthrie, 1967) the research reported here was also designed to answer the following additional questions:

4. Does learning the definition of a veridical concept improve performance on learning to classify examples of that concept?

5. Does having a definition of a veridical concept prior to learning to classify examples of that concept improve performance on subsequent prose learning involving that concept?

METHOD

Subjects

One hundred and thirty-two male and female students from a sophomore level education class at the University of Delaware participated as subjects in the present experiment. Each subject participated individually and was given credit toward his class grade for his participation. Each subject was randomly assigned to 1 of 11 treatment groups.

Design

Table 1 illustrates the design of the present study, specifies the sequence of training, and the number of subjects for each of the 11 treatment groups. It can be seen that for each step in the training sequence a control group is included as a comparison group for the learning data which is obtained at Steps 4, 5, and 7 in the sequence.

Procedures

The training sequence for Group 1 is described here since all of the treatment conditions and the transfer comparisons are based upon the training sequence for Group 1.

As each subject entered the experimental room, he was given an explanation of the purpose of the experiment and a set of printed instructions regarding electrocardiogram (EKG) tracings, the names of the concepts to be learned—ischaemia, infarction, and injury, and a 3 × 5 card with each of the concept names printed upon it.

In Step 1 the subject was taught a conservative focusing strategy for identifying concepts. To ensure that the subjects learned the conservative focusing strategy, classification practice was given and each subject was required to describe aloud just what attribute he was varying with each new card selection.

In Step 2 the subjects then learned to classify stimulus forms of the type used by Bruner in the work on strategy learning. These stimulus forms were presented in a 36-item array (6 × 6). Each subject was given an example card and instructed to determine the general concept of which the card was an example. The subject continued to select examples until he could verbally define the concept correctly (selection paradigm). Each subject continued the classification task until he achieved correct definitions for three different concepts.

Step 3 consisted of training the subjects to identify and label the elements of a normal EKG wave pattern. Each subject was presented with an instruction sheet describing the nature of a normal EKG wave with labels for its critical points, and two pictures of normal EKG patterns—one hand drawn and the other a photocopy of an actual EKG wave. The subject was requested to study the material which was given to him and was permitted to study that material until he signaled

TABLE 1

CLASSIFICATION OF THE TREATMENT GROUPS IN TERMS OF THE TRAINING SEQUENCE

Step	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7
1. Learn classification strategy	X[a]						
2. Classify geometric forms	X	X[a]					
3. Identify and label attributes	X	X	X[a]				
4. Classify EKG schematics							
Inductive	X[b]	X[b]	X[b]	X[b]			
Deductive	X[b]	X[b]	X[b]	X[b]			
5. Test on EKG photocopies	X	X	X	X	X[b]		
6. Study prose passage	X	X	X	X	X	X[b]	
7. Multiple-choice test on prose passage	X	X	X	X	X	X	X[b]

[a] $n = 24$.

[b] $n = 12$.

that the material had been learned. The instruction sheet and the two pictures of the normal EKG patterns were then removed from sight, and the subject was asked to draw and label the critical parts of a normal EKG pattern. The subjects failing to complete the drawing correctly were given an opportunity to review and to draw the pattern again until a perfect drawing was achieved.

In Step 4 of the study the subjects learned to identify (that is, classify) examples of the concepts ischaemia, infarction, and injury. At this point, the group was randomly subdivided into two groups. One of the subgroups learned to classify the concepts inductively (that is, without explanation or description of the characteristics of the concept). The other half of the group learned to classify the concepts deductively (that is, descriptions and explanations of the characteristics of each of the concepts were given at the outset of classification training). In this portion of the experiment a 36-card array of schematic abnormal EKG patterns was used in a selection paradigm. A second 36-card array containing abnormal EKG patterns was available for use if subjects failed to reach a criterion of five correct examples in succession on the first array.

Step 5 of the study consisted of a transfer test in which a series of 30 photocopies of actual EKG tracings was presented to the subjects in a reception paradigm and the number of correct classifications of the 30 tracings was recorded. The subjects were not given feedback as to the correctness of their responses on the transfer test.

Step 6 consisted of a study session. The subjects studied a 2-page prose passage describing the three concepts which they had already learned, some relations among the concepts, and several of the implications of those concepts for medicine. A maximum of 10 minutes was allowed to study the prose passages. All subjects used less than the alloted 10 minutes.

In Step 7 each subject was given a 15-question multiple-choice test on the content of the prose passage. Five of the multiple-choice questions consisted of reading schematics of EKG wave abnormalities. These five items involved the greatest amount of transfer.

RESULTS AND DISCUSSION

The results are presented and discussed in terms of the five questions posed in the introduction.

Question 1: Does either practice at classifying or learning a strategy add to the transfer effects of learning prerequisite content when learning to identify examples of a veridical concept? The answer to this question is based upon the results of comparisons of performance among Groups 1, 2, 3, 4, and 5 on Step 4 (trials to criterion on EKG schematics) and Step 5 (number correct on 30 actual tracings). Group means and variances for performance on Step 4 are shown in Table 2. Analysis of the data obtained in Step 4 shows that for the comparisons of interest none of the groups differed reliably from one another (t ratios in every case are less than one). The results of the analysis indicate that not only did prior practice on classifying or learning a strategy not improve performance over learning prerequisite content, but also that learning prerequisite content (in the form of conceptual attributes) did not improve performance on initial learning to classify the schematic concepts. Said more simply, no form of prior training improved performance on learning to classify schematic exemplars of the concept.

TABLE 2

GROUP MEANS AND VARIANCES ON STEP 4

Group	Trials to criterion on schematics	
	$\bar{X}$	S^2
1	20.1	20.7
2	19.7	21.9
3	18.8	22.3
4	20.3	30.2

Question 2: Does training on identifying and labeling the attributes of a veridical concept improve performance on identifying exemplars of that concept? The analysis of the data obtained from Step 5 bearing on the same question gives a different picture. Inspection of the mean performance for the groups presented in Table 3 shows clearly what the statistical analyses reveal; that is, Groups 1, 2, and 3 do not differ significantly from one another. A comparison of the per-

TABLE 3

MEAN PERFORMANCE OF GROUPS ON STEP 5 (IDENTIFYING ACTUAL EKG TRACINGS)

Learning group	Group				
	1	2	3	4	5
Induction	19.8	18.9	20.4	17.3	
Deduction	22.9	24.5	23.3	21.1	
Totals	21.3	21.3	21.3	19.2	13.9

formance of Group 3 versus that of Group 4 and comparison of Group 4 versus the performance of Group 5 shows that in both cases the mean scores (for Totals) are reliably different ($t = 2.8$, $p < .005$ and $t = 5.4$, $p < .0005$, respectively). The comparison between Groups 4 and 5 reveals that prior experience on classifying schematics facilitates performance on actual EKG tracings. The comparison between Groups 3 and 4, is especially interesting in that these two groups performed equally on the immediately preceeding step involving learning to classify the schematics. Groups 3 and 4 differ on the basis of whether their prior experience included learning to identify the elements of the EKG wave. The analysis shows that although learning to identify the attributes of an EKG wave did not necessarily improve immediate classification performance on the schematics, it did improve performance on subsequent transfer to actual EKG tracing. It may be that prior training on prerequisite content does not necessarily effect improvement on a task where direct practice can be, and is, given but it may improve performance on subsequent tasks where the transfer is greater.

Question 4: Does learning the definition of a veridical concept improve performance on learning to classify examples of that concept? Statistical analysis of the data on Steps 4 and 5 obtained from the inductive and the deductive groups reveals a consistent difference between these two groups in favor of deductive training. Table 3 presents the mean performances for each subdivided group on the transfer test with actual EKG trainings. The statistical comparison of the performance for inductive and deductive subgroups on Step 5 (identifying actual EKG tracings) yielded a statistically significant difference between inductive and deductive groups in favor of deductive training ($F = 38.54$, $df = 1/88$, $p < .0005$). A test of the potential interaction between prior training and inductive–deductive training reveals the difference between groups to be constant (that is, the interaction was not significant). Analysis of the trials-to-criterion data from Step 4 for the same groups reveal the same clear superiority for the deductive learning group ($F = 10.2$, $df = 1/88$, $p < .005$). The results of these comparisons lend strong support to the notion that learning to classify examples of a veridical concept proceeds most rapidly when a clear exposition of conceptual attributes is supplied.

TABLE 4

Mean Number Correct on Prose Test for All Experimental Groups (Summed across Induction–Deduction)

Statistic	Group						
	1	2	3	4	5	6	7
$\bar{X}$	12.2	11.5	11.9	9.9	8.6	10.3	5.6
S^2	2.3	4.6	2.4	2.9	1.7	2.5	1.7

TABLE 5

Summary Table of t Ratio and p Value for Comparisons of Interest

Comparisons	t ratio
Learning attributes and classfying schematics (Groups 1, 2, 3, 4 vs. 5, 6)	4.97**
Learning attributes (Groups 1, 2, 3, vs. 4)	3.69**
Classifying schematics (Group 4 vs. 5, 6)	.91*
Studying passage (Groups 5, 6, vs. 7)	7.66**

* $p < .50$.
** $p < .005$.

Answers to the remaining questions are based upon data obtained in Step 7 of the training sequence (a multiple-choice test on the prose passage).

Question 3: Does prerequisite concept learning improve subsequent prose learning involving those concepts? Table 4 contains the mean number correct on the final test for all experimental groups summed over induction and deduction. Table 5 contains the summary of the results of statistical comparisons among the treatment groups which are of primary interest. It is evident from Table 5 that those groups having an opportunity to learn the attributes of a concept prior to learning to classify instances of the schematics (Groups 1, 2, and 3) performed better on the prose test than did the group which only classified instances of the schematics without prior learning of the attributes (Group 4). This finding is consistent with the earlier finding where prior learning of the attributes improved performance on subsequent tests of performance on classifying actual EKG tracings.

It is worthy of note that on the prose test, Group 4 which had learned to classify to criterion on the EKG schematics still did not perform better than Groups 5 and 6 which could not profit from prior learning. In the present study, at least, one would have done as well simply by studying the prose passage, and not having learned to classify examples of the concept. Learning to classify conceptual instances without labeling the attributes appears to be of little transfer value in subsequent prose learning. This consistent facilitation of subsequent performance as a result of prerequisite learning of conceptual attributes emphasizes the potential value of content-specific transfer for subject-matter learning. It is important to note facilitation on the prose test occured only on items which required performance on schematic drawings of EKG waves (see description of Step 7). On items which could be answered solely on the basis of the preceding prose passage there were no significant differences between any of the treatment groups, with the exception, of course, of Group 7 which performed significantly poorer.

Question 5: Does having a definition of a veridical concept prior to learning to classify examples of that concept result in improved performance on subsequent prose learning involving the concept? The mean performance on the prose test for the inductive–deductive learning groups is presented in Table 6.

TABLE 6

MEAN NUMBER CORRECT ON PROSE TEST FOR INDUCTIVE-DEDUCTIVE LEARNING GROUPS

Learning group	Groups with pretraining			
	1	2	3	4
Induction	12.7	10.4	11.6	9.1
Deduction	11.8	12.6	12.3	10.7

As in the results of the analysis of performance on reading actual EKG tracings, the deductive learning groups generally performed better on the final test, than did the inductive learning groups (overall $F = 3.63$, $df = 3/88$, $p < .05$). A careful inspection of the test performance revealed that the better performance by inductive learners in Group 1 was primarily attributable to a disproportionately greater number of the factual items being answered correctly by this group (that is, those items which did not require reading of EKG tracings). It is difficult to develop a sensible explanation for this inconsistency. The fact that Group 1 performance on all other analyses was consistent with Groups 2 and 3 suggests that the obtained significant interaction was spurious.

In sum, the results of the present study suggest the following conclusions:

1. Classification practice and learning a selection strategy does not readily produce significant improvement in learning complex subject-matter concepts.

2. Prior training on the elements (learning to identify and label attributes) of a subject-matter concept is a potent variable influencing subsequent performance on transfer tasks involving those concepts.

3. Deductive training is more efficient when learning to identify conceptual instances and contributes to improved performance on subsequent prose learning involving those concepts.

4. Direct practice on classifying instances of a subject-matter concept improves performance on the practice instances but seems to have little transfer value.

The conclusions from the present study would seem to suggest that educational practitioners might profitably arrange curricula by programming for content-specific transfer and deductive learning when possible.

REFERENCES

BRUNER, J. S., GOODNOW, J., & AUSTIN, G. A., *A study of thinking*. New York: Wiley, 1956.

CARROLL, J. B. Words, meanings and concepts. *Harvard Educational Review*, 1964, **34**, 178–202.

DECECCO, J. P. *The psychology of learning and in-*

struction: Educational psychology. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

GAGNÉ, R. N. *The conditions of learning.* New York: Holt, Rinehart & Winston, 1965.

GUTHRIE, J. T. Expository instruction versus a discovery method. *Journal of Educational Psychology,* 1967, **58,** 45–49.

JENSEN, A. R. Individual differences in learning: Interference factor. (United States Office of Education Cooperative Research Project No. 1967) Washington, D. C.: United States Government Printing Office, 1965.

SUPPES, P. Mathematical concept formation in children. *American Psychologist,* 1966, **21,** 139–150.

(Received May 11, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 5, 371-375

EFFECT OF INCENTIVE VARIABLES AND TYPE OF ADJUNCT QUESTION UPON TEXT LEARNING

LAWRENCE T. FRASE[1]
Bell Telephone Laboratories, Murray Hill, New Jersey

This study explored the effects of factual and inferential adjunct questions upon text learning and how their effects are altered by incentives. Telling college students, before they read, that monetary rewards would be given or not given for posttest performance reduced errors on adjunct questions, improved recall for facts and the adjunct questions, and enhanced the facilitative effects of inferential adjunct questions. Informing the subjects of the incentive conditions immediately before testing had relatively little effect. Subjects recalled mainly the facts in the text, regardless of the type of adjunct question they had seen.

One purpose of this study was to determine whether subjects are as poor at remembering inferential material as research suggests (Frase, 1969; 1970a). For example, in the latter study, the subjects evaluated all of the inferences that were assessed on a posttest, yet they recalled only 15% of them. Recall for facts in the text was 35%. The work of Bormuth (1968) and Frase and Washington (1970) with children confirms that human ability to put together information from different sentences can be surprisingly low. If the lack of ability to combine text information is generally supported, it could have useful implications for understanding the cognitive outcomes of reading, for complex verbal structure depends upon the elaboration of remote connections among facts in a text.

Another purpose of this study was to confirm that inferential questions produce relatively high recall of factual material (Frase, 1969). Factual questions evidently restrict the range of relevant text sentences, whereas inferential questions increase the number of relevant sentences and perhaps ensure the repeated processing of those sentences, with a concomitant increase in time spent reading (Frase, 1970b).

Another purpose of this study was to determine how the effects of inferential and factual adjunct questions are modified by the incentive under which reading occurs. A study by Frase, Patrick, and Schumer (1970) showed that time and amount learned increased with the amount of money the subjects expected to be paid for posttest performance. The question arises whether this effect is only a consequence of trying harder on the posttest. The results of Ausubel, Schpoont, and Cukier (1957) indicate that intentions to remember that arise after exposure to a text do not facilitate retention. If the behaviors that produce learning occur primarily during reading, then informing subjects of the incentive conditions before reading should produce higher recall than informing them immediately after reading (but before testing). As in the previous study (Frase et al., 1970), subjects who perform with hopes of monetary reward should recall more than those who expect not to be rewarded.

The addition of incentives should increase the tendency to make effective use of the additional cues provided by inferential questions. Incentive should thus enhance the facilitative effects of inferential questions, especially if knowledge of the incentive conditions is available before reading (since the most potent learning behaviors are assumed to occur during exposure to the text and not during testing).

METHOD

Subjects

Forty-eight volunteer college undergraduates from Montclair State Teachers College, New Jersey, were paid for participating.

---

[1] Requests for reprints should be sent to Lawrance T. Frase, Bell Telephone Laboratories, Mountain Avenue, Murray Hill, New Jersey 07974.

Materials

The texts consisted of three passages used in other experiments (Frase, 1969). One passage described people of a foreign country, another described the production of automobiles, and the last described astronomical discoveries. Each passage consisted of four experimental text sentences, embedded in other material. The experimental sentences may be represented as: "As are Bs," "Bs are Cs," "Cs are Ds," "Ds are Es." Each letter stands for a logical class. The sentences in each text were given in the order indicated. The following is an example of one passage with the logical classes italicized:

> Scientists consider all known *planets within fifteen light years* distance to be *interesting*. The only *interesting* planets are the *"Blue" planets*; however later discoveries may change this judgment. There is general agreement on the fact that all the *"Blue" planets* are in *Galaxy IV*. It is hoped that new discoveries will be made in the near future which will extend this knowledge. Recently some new facts have been discovered about our universe. The planets in *Galaxy IV* are capable of *supporting life*, according to all present evidence.

Adjunct questions. Two adjunct questions were typed above each passage. The questions were in the form of a sentence. The subject was to determine whether the conclusion could be drawn from the passage directly below it, and he responded by circling one of two response options, "valid" or "invalid." *Factual* questions were of the form, "As are Bs" and "Cs are Ds." *Inferential* questions were of the form "As are Ds" and "Bs are Es." For both factual and inferential questions four different logical classes were relevant in the passage, but an inferential question required combining three text sentences to evaluate the problem conclusion. All problem sentences were valid conclusions from the texts. In the total set of questions, 75% of the logical classes were the same for factual and inferential adjunct questions. Each subject saw the same type of question on each of the three passages, that is, either inferential or factual questions.

Incentive level. Half of the subjects (paid) were informed that they could earn additional money up to a total of $4.20 for the experiment if they could answer the questions on the posttest. The other half of the subjects saw the incentive instructions (as in Frase et al., 1970), but the instructions informed them that they were in the not paid group. All of the subjects were informed before reading that a test would follow.

Incentive position. The incentive information was placed immediately before the reading passages, or immediately after the reading passages under the cover sheet of the test booklet.

Test. Immediately after reading all three passages, the subjects turned to the test booklet in which they were instructed to write everything that they could recall about each text, including all the valid inferences they could possibly generate. There was a separate page for each text, given in the same order as the passages had occurred. Free recall for facts and inferences was scored by determining whether the subject had asserted a relationship of the form "As are Bs," etc. A more extensive discussion of the scoring procedure is described in Frase (1969). Two text sentences and two inferences (which differed from the adjunct questions used with the texts) were selected from each of the three passages, for a total possible score of six facts and six inferences for each subject. The same logical classes were involved in these incidental facts and inferences. For instance, if the facts were "As are Bs; Cs are Ds," then the inferences tested would be "As are Cs; Bs are Ds." In all cases, the inferential test sentences only involved putting together two text sentences to draw an inference.

Procedure

The subjects were randomly assigned to the eight experimental conditions, and they were run in small groups. Each subject received separate booklets containing the passages and the tests,[2] each with their associated instructions (including incentive information where appropriate). Instructions before the passages informed all subjects that there would be a test covering the content of the passages after reading, and that

> . . . our objective in having you solve problems is to aid you in learning the information. We have found that people learn more from solving these problems than they do if they study the passages without the problems. You have as much time as you need, so work carefully.

After completing each page of the text or test material, the subject was to go on to the next page without going back. There was no time limit on reading or testing.

Design

The design was a 2 × 2 × 2 × 2 analysis of variance, with repeated measures on the last factor. The factors were type of adjunct question, incentive level, incentive position, and type of recall (fact or inference). A separate analysis was carried out on recall for assertions incidental to the adjunct questions, and another on recall for the adjunct questions. A 2 × 2 × 2 analysis was also conducted on correct solutions to the adjunct questions.

---

[2] A recognition test, including all inferences and facts embedded in distractors, was administered after the free-recall tests. The results replicate those of the recall tests, and hence have been eliminated to simplify the discussion. The one notable exception, for which a rationale has been given by Frase (1970a), was that the recognition tests did not reveal a difference between recall of facts and inferences. The rationale for this finding concerns the fact that the recognition test explicitly combined all pairs of logical classes, thus eliminating a significant cognitive process involved in recall.

RESULTS

Performance on Adjunct Questions

Although the Inference group scored somewhat lower on solutions to the adjunct questions ($\bar{X} = 83.2\%$) than the Fact group ($\bar{X} = 90.8\%$), the difference was not statistically significant ($F = 3.5$, $df = 1/40$, $p < .1$). Incentive information before reading ($\bar{X} = 91.5\%$) produced more correct solutions on the adjunct questions than when it was given after reading $\bar{X} = 82.4\%$; $F = 5.0$, $df = 1/40$, $p < .05$). Pay produced more correct solutions on the adjunction questions ($\bar{X} = 93.0\%$) than no pay ($\bar{X} = 81.0\%$; $F = 8.6$, $df = 1/40$, $p < .01$).

Incidental Recall

Free recall of the facts and inferences incidental to the adjunct questions revealed, as in previous studies (Frase, 1969, 1970a), that inferential adjunct questions produced higher recall ($\bar{X} = 21.2\%$) than factual questions ($\bar{X} = 5.6\%$; $F = 37.7$, $df = 1/40$, $p < .001$).

Incentive information before reading resulted in higher recall ($\bar{X} = 21.2\%$) than after reading ($\bar{X} = 10.8\%$; $F = 9.3$, $df = 1/40$, $p < .005$). Incentive position also interacted with type of adjunct question ($F = 9.2$, $df = 1/40$, $p < .005$). The Fact group recalled 5.6% of the facts and inferences whether incentive information was given before or after reading. The Inference group recalled 36.7% of the facts and inferences if incentive information was given before reading, but only 16.1% if incentive information was given after reading. Overall, incentive level was not significant.

Recall for facts was 22.6%, for inferences it was 9.4% ($F = 30.7$, $df = 1/40$, $p < .001$). The type of question seen in the text interacted with the incidental information that the subjects recalled ($F = 4.14$, $df = 1/40$, $p < .05$). Factual questions produced 9.8% recall of the incidental factual material whereas inferential questions produced 35.4% recall of those facts. For recall of incidental inferences, the mean for the Inference group was 17.3%, and for the Fact group it was 1.4%. The interaction between type of question and whether subjects remembered factual or inferential content may be due, in part, to the low overall scores for the Fact group.

Recall of Adjunct Questions

Interesting data come from an analysis of recall for the statements used as adjunct questions.

Incentive information given before reading produced higher recall of the adjunct questions ($\bar{X} = 31.8\%$) than when it was given after reading ($\bar{X} = 21.5\%$; $F = 6.76$, $df = 1/40$, $p < .025$). There was no overall difference between the Fact and Inference groups, nor between incentive levels.

Recall for the factual questions was 45.6%, and for inferential questions it was 7.7% ($F = 91.04$, $df = 1/40$, $p < .001$). The expectation was that the subjects in the Fact group would recall the facts they had seen as adjunct questions, and the Inference subjects would remember the inferences they had seen. This interaction between type of question and level of recall for the inferential and factual questions was borne out by the data ($F = 5.24$, $df = 1/40$, $p < .05$), but the Inference subjects did not remember any more inferential questions (14.0%) than they did incidental inferences (17.3%). Recall for inferential question statements by the Fact group (which had not seen those statements) was 1.4%. Recall for factual questions was 48.5% for the Fact group, and 42.5% for the Inference group (which had not seen those sentences as questions). In short, recall for the inferential adjunct questions was not greatly facilitated for subjects in the Inference group, but they remembered the facts nearly as well as subjects tested explicitly over those facts. Regardless of the type of adjunct question, the primary content recalled by all subjects (including the questions and incidental material) was simple factual material.

DISCUSSION AND SUMMARY

The present study showed that inferential questions, which have apparently little direct instructive effects, can produce as high incidental factual recall as questions that explicitly test for those facts. Enhanced recall with inferential adjunct questions presumably occurs because such questions subsume factual material, thus insuring that subjects respond actively to the text information.

The problem arises as to why recall for the inferential adjunct questions was not facilitated to a greater extent, for the Inference subjects did not recall the inferential statements they had seen as adjunct questions any better than incidental inferences. Nominally, the problem statements had the same status as the factual material; they were an explicit statement of a relationship. But inferential problems differ cognitively from factual content in the following ways: (a) there is some uncertainty about their truth value, (b) they can be stored *temporarily* to control logical processing, unlike factual content which must be operated upon to solve the problem, and (c) they are followed by increased learning activities. There is no way to decide upon the most important of these three alternatives from the present data. According to the third alternative, an inferential question produces activities which lead to the recall of several text sentences; hence there is more in memory to interfere with the recall of the question itself. Such interference might diminish if the text were learned more completely. A previous study (Frase, 1969) indicated that when an inferential question required two sentences to answer, the probability of recalling a fact or inference was .60, whereas the probability of recalling the question was .25. When the four experimental sentences were required to answer the adjunct question, the probability of recalling a fact or inference was .75, and the probability of recalling the question was .55. Although these data are not very convincing for the interference interpretation, the possibility is an intriguing one that deserves further exploration. The suggestion is that the learning consequences of an instructional aid may react differentially upon various components of the instructional episode. In any case, the present results indicate that the inferential adjunct statements had no special significance in memory.

As predicted, recall for incidental inferences was quite low, although subjects were able to solve such problems well with the text present. The findings of this study agree with the data from an earlier study (Frase, 1970a) in this respect.

Incentive evidently made the subjects work harder at the task; errors on the adjunct questions were reduced if subjects expected monetary reward. Knowing that rewards were to be given (even to other people) enhanced the positive effects of the inferential questions.

Although incentive information was clearly most useful when it preceded reading, the fact that incentive level did not interact significantly with the position of incentive information suggests that the competitive learning situation may have facilitated performance. For instance, both Incentive-Before groups (whether they were to be paid or not paid) were superior to their corresponding Incentive-After groups on recall of the adjunct and incidental information. Similar results were obtained for correct solutions to the adjunct questions. Although overall learning differences produced by incentive in a previous study (Frase et al., 1970) were slight, they were significant. The present findings do not confirm those of Frase et al. (1970), although the format for presenting the incentive information was similar in the two studies. One possible reason for the lack of incentive effect on recall could be that the subjects in the present study were volunteers who were not required to participate as a course requirement. Given that these subjects were volunteers, their initial level of motivation might have been relatively high.

REFERENCES

AUSUBEL, D. P., SCHPOONT, S. H., & CUKIER, L. The influence of intention on the retention of school materials. *Journal of Educational Psychology,* 1957, **48,** 87–92.

BORMUTH, J. R. The effectiveness of current procedures for teaching reading comprehension. Paper presented at the meeting of the National Council of Teachers of English, Milwaukee, November 1968.

FRASE, L. T. Structural analysis of the knowledge that results from thinking about text. *Journal of Educational Psychology Monograph Supplement*, 1969, **60,** (6, Pt. 2).

FRASE, L. T. Influence of sentence order and amount of higher level text processing upon reproductive and productive memory. *American Educational Research Journal,* 1970, **7,** 307–319. (a)

FRASE, L. T. Structure and learning from written materials. In Arthur M. Taylor (Chm.), *Educational implications of research on elaboration, imagery, and memory.* Symposium presented at the annual meeting of the American Educational Research Association, Minneapolis, March 1970. (b)

FRASE, L. T., PATRICK, E., & SCHUMER, H. Effect of question position and frequency upon learning from text under different levels of incentive. *Journal of Educational Psychology,* 1970, **71,** 52–56.

FRASE, L. T., & WASHINGTON, E. D. Children's ability to comprehend text. Paper presented at the meeting of the American Psychological Association, Miami, September 1970.

(Received May 29, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 5, 376-383

ELABORATION TRAINING AND PAIRED-ASSOCIATE LEARNING EFFICIENCY IN CHILDREN[1]

WILLIAM D. ROHWER, JR.[2] AND MARY SUE AMMON[3]
University of California, Berkeley

Samples of 60 high-SES white and 60 low-SES black second-grade children were assigned in equal numbers to each of three treatment conditions: training, practice, and control. In the training condition, subjects were instructed in verbal and visual methods of elaborating noun pairs to increase their efficiency in learning paired associates. Treatment effects were assessed in terms of performance on two posttest paired-associate tasks: an aural list consisting of 25 noun pairs presented aurally, and a mixed list, consisting of 25 noun pairs presented both aurally and visually in five different ways. The results revealed a positive effect of elaboration training for both samples but the effect was confined to the aural list for the high-SES whites and to the mixed-list for the low-SES blacks.

The present study was designed to evaluate an attempt to increase, through training, paired-associate learning efficiency in children. Such an attempt is of interest from two perspectives—one psychological, the other educational. In the psychological domain, several recent experiments indicate that substantial increments in paired-associate learning efficiency can be produced by the elaboration of the pairs to be learned. The specific meanings of the term elaboration can be illustrated in connection with a particular task, that of learning a list of noun pairs. In this kind of task, elaboration may be directed at either the individual items that comprise each pair or at the pair unit itself. Consider first the case of elaborating individual items. If the materials are presented aurally, elaboration consists of representing the referents of each noun visually as in a picture or, more loosely speaking, in an image; if the materials are presented pictorially, elaboration consists of representing each item verbally as by the appropriate noun label for the given object. Second, consider the case where both members of each pair are included in a single elaborative unit. As in single-item elaboration, pair elaboration may be either verbal or pictorial: the two nouns can be used to form the subject and object of a sentence describing an event; or, an event involving the two objects named by the nouns can be depicted pictorially.

Several strands of evidence support the presumption that each of these four forms of elaboration can increase the efficiency of learning noun pairs. With respect to single-item elaboration, it has been demonstrated repeatedly (cf. Paivio, 1969) that the learning of high-imagery-value noun pairs is more efficient than the learning of low-imagery-value noun pairs. Similarly, Rohwer, Ammon, Suzuki, and Levin (1971) have reported that paired-associate performance is better when noun labels are presented concurrently with pictorially presented object pairs than when the pictures are presented alone.

The positive effect of pair elaboration on learning efficiency has been documented by means of three different methodologies: postlearning interviews, prelearning instructions to elaborate, and manipulation of the conditions of presentation. The postlearning interview technique has revealed (*a*) that subjects report elaborating paired associates by constructing sentences containing the word pairs or by forming images involving the referents of the pairs (e.g., Bugelski, 1962; Runquist & Farley, 1964); (*b*) that the kinds of elaboration reported can be classified reliably with respect to complexity (Martin, 1967); and, (*c*) that there is a positive correlation between complexity of the elaboration reported and the efficiency of paired-associate learning (Martin, 1967; Montague & Wearing, 1967). The efficiency of paired-associate learning has also been shown to vary as a function of whether or not the pairs are presented in an elaborated form. In this connection, both verbal and visual forms of elaboration have proved to be effective; performance is increased by presenting noun pairs in the context of sentences or relational phrases (Davidson & Adams, 1970; Rohwer, 1967) or by depicting the referents of the nouns in a pictorial interaction (Milgram, 1967b; Reese, 1965; Rohwer, Lynch, Levin, & Suzuki, 1967). Finally, instructions to elaborate noun pairs either by forming sentences or images have been observed to produce substantial increments in performance (Bower, 1969; Jensen & Rohwer, 1963, 1965; Milgram, 1967a).

---

[1] This work was supported in part by a contract with the United States Office of Economic Opportunity (OEO 2404). The report was prepared at the Institute of Human Learning, University of California, Berkeley, which is supported by grants from the National Institutes of Health.

[2] Requests for reprints should be sent to William D. Rohwer, Jr., Institute of Human Learning, University of California, Berkeley, California 94720.

[3] We wish to acknowledge the indispensable assistance of the teachers and children in those schools where the data were collected as well as the contributions of Nancy Suzuki in administering the various tests and of Joel R. Levin in analyzing the results.

Since brief elaboration instructions have demonstrable effects on paired-associate performance when given immediately before the onset of the learning task, more extensive training in elaboration skills should produce more enduring effects. It was expected that such training would result in detectable positive transfer in performance on paired-associate lists administered outside of the context in which the training was provided. Thus, the primary question of psychological interest was whether or not elaboration training can be shown to make durable differences in the efficiency with which children learn paired-associate lists.

From an educational perspective, two features of the present study were prominent. The first was whether elaboration training improves paired-associate performance more than simple practice on paired-associate tasks for equal amounts of time. The second concerned the issue of whether or not elaboration training would suffice to reduce observed discrepancies in learning efficiency between children classified as low-SES (socioeconomic status) black and children classified as high-SES white. Although the differences between these two populations are much smaller on paired-associate tasks than on intelligence and achievement tests, they are frequently detected among young children, that is, 8 years of age and under. In particular, they are detected when the method of presenting paired-associates does not provide *item* elaboration and when it does provide *pair* elaboration (Rohwer et al., 1971). Accordingly, the paired-associate tasks used to evaluate the effects of elaboration training were constructed to permit an evaluation of this interaction between populations and methods of presentation.

METHOD

Subjects

Sixty children were randomly selected from the second-grade classes in each of two public elementary schools. One of the schools serves a low-SES black residential area and the other serves a high-SES white area. The SES designations were based on average census tract information collected in the 1960 survey.

Design

The factors in the 2 × 3 × 20 design were populations (high-SES white versus low-SES black), treatments (training, practice, control), and levels. Assignment to levels was determined by pretest performance; within each population, the sample was subdivided into 20 groups of three subjects each in terms of pretest similarity and the members of each level were randomly assigned to treatment conditions.

The three treatments were distinguished by the activities prescribed for the pretest–posttest interval. In the training condition, instruction was designed to facilitate the acquisition of the following skills in connection with paired-associate learning tasks: (*a*) self-repetition of presented pairs; (*b*) verbal labeling of objects presented pictorially; (*c*) visualization of noun referents presented aurally; (*d*) the visualization of action sequences involving object pairs; and (*e*) the generation of sentence descriptions of such action sequences. Instruction relating to these skills was conducted in connection with several paired-associate lists comprised of a variety of types of items—objects, pictures, aurally presented nouns. The following are examples of training activities. The trainer presented an array of three-dimensional toy objects and asked the child to select any two and to perform an action episode involving them. The child was then asked to describe the performance or "story" verbally. In another activity, pictures of object pairs were presented on a television monitor; the child was directed to visualize the picture story involving the two objects and to give a verbal description of the story or to guide the trainer in making a drawing of the episode.[4]

Instruction in the practice condition was directed only at the skill of self-repetition of presented pairs. Thus, the major activity in the training sessions for this condition was the learning of successive paired-associate lists.

Certain other features of these two treatment conditions were common to both. In both content and presentation mode, the paired-associate lists used for instructional purposes in the training condition were identical with those used in the practice condition. A variety of presumed incentives were made available to children in both conditions: a chart was made for each child which displayed a record of performance improvement as a function of sessions, colorful stickers were awarded for successful performance, and the child was allowed to play with hand puppets after performing on a paired-associate list. In sum, children in both conditions received equivalent exposure to the same kind and number of paired associates and received the same kinds of feedback in connection with their performance. Thus, the structure of the practice condition was designed to control for the effects of all experiences provided in the training condition except those involving direct instruction in elaboration techniques.

In contrast, children assigned to the control condition received only the pretest and the posttest.

Materials

Both the pretest and the posttest consisted of two different kinds of 25-item lists of noun pairs. In each case the two lists differed with respect to the manner in which the items were presented: in the one (aural) all pairs and test stimuli were presented aurally by means of an audio tape recorder; in the other (mixed), equal numbers of items were presented in each of five different ways. Thus, the mixed paired-associate list yielded scores on five item types: noun pairs presented aurally (nouns); photographs of object pairs depicting noun referents (pictures); photographs of similar objects with their noun labels presented aurally (nouns–pictures); photographs of object pairs with their noun labels presented aurally in the context of a sentence (sentences–pictures); and photographs of an event involving the members of object pairs with their noun labels presented aurally (nouns–action). For the mixed list, the test stimuli consisted of one of the items from each pair presented in the same manner as on the study trials. For example, test stimuli for pairs of the pictures type were photographs of single objects from each of the relevant pairs. All materials for the mixed list tests were recorded on videotape and presented by means of a television monitor.

For both the mixed and the aural lists, the pre- and posttests were administered for two study and two test trials each. In the case of both kinds of lists, the rate of presentation on study and on test trials was 4½ seconds and the intertrial intervals were 4½ seconds. The order of items was different on every trial.

Two considerations led to the choice of the particular paired-associate lists used as the pre- and posttests. The first pertains to the mixed lists which were selected to permit an assessment of the differential effects of training as a function of presentation methods. Furthermore, these lists had been used previously in such a way as to yield both reliability and difficulty estimates for children from populations similar to those sampled in the present study. The consideration leading to the use of the aural lists was that of presumed sensitivity to training effects. A list of 25 aurally presented paired associates is more difficult than a mixed list of comparable length. Accordingly, it was reasoned that the more difficult aural list might detect training effects that would not appear in the easier mixed list.

Procedure

The study was conducted by two female experimenters, one of whom administered all pre- and posttests while the other administered all of the treatment sessions. The experimenter responsible for the pre- and posttests did not participate in the assignment of subjects to treatment conditions nor was this information available to her at any time during the study.

The major events of the study consisted of the administration of the two pretest lists, the assignment of subjects to levels, the assignment of subjects within levels to treatment conditions, the administration of the various treatments, and the administration of the posttest lists.

The two pretest lists were given to the subjects individually. The order of administration was constant across all subjects: the mixed list followed by the aural list. Prior to the onset of each list, instructions were given informing the subject of the nature of the materials to be presented and of the task he was expected to perform. These instructions were illustrated by five sample items presented in accord with the same study-test pattern as that which characterized the test lists. The sample items were repeated until the subject responded correctly to three of the five test stimuli in the list. Following the attainment of this criterion, the subject was told that the learning task proper would begin and the 25-item list was administered.

For the purpose of assignment to levels and to treatment conditions, two scores were calculated for each subject: total number of correct responses on the two test trials of the mixed list, and total number of correct responses on the two test trials of the aural list. Within each of the two populations sampled, the subjects were then grouped into sets of three such that the members of each set were characterized by similar scores on both the mixed and the aural lists. Thus, subjects having high scores on both lists were grouped together, those having low scores on both were grouped together and those having high scores on the mixed and low scores on the aural list were grouped together, and vice versa. This procedure yielded 20 levels of three subjects each. Within each level, subjects were randomly assigned to the three treatment conditions.

---

[4] A detailed script used in conducting the training sessions may be obtained from the first author on request.

The method of assigning the subjects to levels in terms of their scores on the two pretest lists considered separately permitted the application of a randomized analysis of variance design to the results yielded by each of the two posttest lists. It should be noted that the same rank order of levels did not obtain across the analyses for the two posttests. For example, the three subjects assigned to the highest ranking level for the mixed list analysis may have occupied a somewhat lower ranking level for the analysis applied to the results of the aural list. Nevertheless, the various sets of three subjects remained intact across both analyses even though the rank of the sets among the levels varied from one analysis to another.

The training and practice treatments were administered in five sessions of 20 to 30 minutes duration each on five successive school days. Two subjects were given the assigned treatment sessions together rather than singly in order to conserve time. The treatment sessions for the high-SES white sample were conducted during the first and third weeks of this phase of the study and those for the low-SES black sample were conducted during the second and fourth weeks. For each subject a record was made of the number and the identity of the treatment sessions missed in order to permit a post assessment of the relationship between attendance and posttest performance. In all cases, sessions missed were attributable to absence from school.

The schedule developed for the posttest phase specified that equal numbers of subjects from each of the three treatment conditions would be administered the final mixed and aural lists during the week immediately following the treatment sessions for those assigned to the training and practice conditions. This schedule was followed insofar as possible; departures from it were due to school absence. In such cases, the posttest was administered upon the subject's return to school. A record was made of the number of days elapsing between the end of the treatment sessions and the administration of the posttest. The procedure followed in giving the posttest lists was the same as that described for the pretest lists.

TABLE 1

MEAN NUMBERS OF CORRECT RESPONSES ON THE AURAL-LIST AND MIXED-LIST POSTTESTS AS A FUNCTION OF POPULATIONS AND TREATMENTS

Populations	Training	Practice	Control	Total
Aural list				
High-SES white	5.95	4.68	3.92	4.85
Low-SES black	2.68	2.40	2.58	2.55
Mixed list				
High-SES white	14.02	14.15	12.18	13.45
Low-SES black	12.98	11.58	11.62	12.05

Note.—The maximum score for both lists = 25; *MSE* = 11.51 for the aural list; 13.95 for the mixed list; $df = 74$.

RESULTS

Pretest performance was used to assign subjects to levels within populations. Accordingly, the effects of the treatment conditions were assessed entirely in terms of performance on the posttests, aural and mixed list.

Aural Posttest

The results yielded by the aural posttest are shown in Table 1 as a function of populations and treatments. The data were evaluated by analysis of variance in which the sources assessed were: populations, levels nested within populations, and the simple main effects of treatments within populations; the variance associated with treatments was partitioned into components to yield two planned comparisons: practice with control, and the average of practice plus control with training. In the case of the latter comparison, a directional prediction favoring the training condition was made in advance thus calling for a one-tailed test; the remaining tests were two-tailed and all were made with $\alpha = 05$.

As the means in Table 1 suggest, the populations difference was significant ($F = 27.58$, $df = 1/74$, $p < .01$)[5], the high-SES white subjects scored higher than the low-SES black subjects. The main effect of levels was significant for the high-SES whites ($F = 3.13$, $df = 19/74$, $p < .01$) but not for the low-SES blacks ($F < 1$). The difference in the magnitude of these two F ratios may indicate that performance on the aural list was more stable between the pre- and posttests for the high-SES whites than for the low-SES blacks.

---

[5] In all analyses reported, observations were estimated for two subjects lost by virtue of absence from four of five training sessions and during the entire posttest period. Therefore, all relevant *df*s have been adjusted accordingly.

With regard to the question of principal interest, the analysis of variance revealed quite different effects of treatments for the two populations. For the high-SES whites, the difference between the practice and control conditions was not significant ($F < 1$) whereas performance in the training condition was superior to the average of those two conditions ($F = 6.31$, $df = 1/74$, $p < .01$). Practice on the task of learning paired-associate lists had no detectable effect on posttest performance on the aural list but the instruction provided in the training condition was successful in improving performance on this same criterion test. The results for the low-SES black samples were not at all comparable: the F ratios for both planned comparisons were less than 1. Neither practice in paired-associate learning nor specific training in elaboration skills resulted in performance increments detectable on the aural posttest.

Mixed-List Posttest

The results of the study as indexed by the mixed-list posttest are presented in Table 1 as a function of populations and treatments. For the between-subjects factors, the analysis design applied to the mixed-list posttest was the same as that described for the aural posttest. The main effect of populations was significant once again ($F = 8.33$, $df = 1/74$, $p < .01$) with more correct responses observed for high-SES whites than for low-SES blacks. The variance associated with levels was significant in both populations (Fs = 2.96 and 2.74; $df = 19/74$; $p < .01$, for the high-SES whites and the low-SES blacks, respectively). Thus, in contrast to the outcome on the aural posttest, performance seems to have been of equivalent degrees of stability for both populations on the mixed-list posttest.

An inspection of Table 1 suggests and the analysis confirms that performance in the practice condition for the high-SES whites was significantly superior to that in the control condition ($F = 5.59$, $df = 1/74$, $p < .05$). Indeed, descriptively speaking, the training and practice conditions differ only negligibly and both may be presumed superior to the control condition. Accordingly, it may be inferred that for this sample, training had no specific effects beyond those produced by simple practice on paired-associate tasks. It is consistent with this interpretation that the variance associated with the comparison between the training condition and the average of the practice and control conditions was not significant ($F = 1.42$, $df = 1/74$, $p > .05$). These results are in marked contrast to those observed for the low-SES black sample. For this population, the practice condition afforded no advantage over the control ($F < 1$), but the training condition was superior to the average of practice and control ($F = 3.61$, $df = 1/74$, $p < .05$).

In summary, the analyses of between-subjects sources of variance reveal that the effects of training depend on both the character of the posttest list and the population sampled; the training condition produced specific facilitative effects for the high-SES whites only on the aural posttest, not on the mixed-list posttest, whereas for the low-SES blacks, the training condition facilitated learning on the mixed-list posttest but not on the aural list.

It will be recalled that the mixed-list posttest was comprised of five different types of items: nouns, pictures, nouns–pictures, sentences–pictures, and nouns–action. For the multivariate analysis designed to estimate the amount of variance attributable to these differences among methods of presenting the various paired associates, four orthogonal linear transformations of the five original item type variates were made. The effect of item types was significant ($F = 391.63$, $df = 4/71$, $p < .01$). Another multivariate test of the transformed variables revealed that the effect of populations was not the same across all of the variates ($F = 5.36$, $df = 4/71$, $p < .01$). Differences among the original item types variables that are relevant to these tests are displayed in Table 2 as a function of populations.

An inspection of Table 2 suggests that the populations difference may have been confined to the item types, nouns–pictures, sentences–pictures, and nouns–action; descriptively, the differences on the nouns item type and the pictures item type appear negligible. Since this pattern of results across item types varied substantially from that reported previously (Rohwer et al., 1971), an examination was made of performance on the pretest. The relevant results are also presented in Table 2 as a function of populations and item types.

It must be emphasized that the pretest analysis included the scores of subjects later lost from the training study for absenteeism so that there are no meaningful direct comparisons between the results for the two tests. Nevertheless, the difference in the pattern of correct responses across item types is interesting in three respects. First, note that the item type showing the greatest populations difference on the posttest, namely, nouns–pictures, revealed virtually no populations difference on the pretest. Second, on the pretest, the two populations appeared to differ on all of the remaining four item types, including those showing virtually no difference on the posttest, that is, nouns and especially pictures. Finally, note that the nouns–pictures item type on the pretest was associated with more correct responses than the pictures type, pronouncedly so for the low-SES blacks, whereas on the posttest, the pictures item type was superior to nouns–pictures for both populations. To be sure, this inconsistency between pre- and posttest performance may be attributable to list differences. The fact, however, that both these lists, the pretest list and the posttest list, have produced quite consistent patterns of results in previous use (Rohwer et al., 1971) does not encourage adoption of this interpretation.

TABLE 2

MEAN NUMBERS OF CORRECT RESPONSES ON THE MIXED-LIST POSTTEST AND PRETEST AS A FUNCTION OF POPULATIONS AND ITEM TYPES

Populations	Item types				
	Nouns	Pictures	Nouns–Pictures	Sentences–Pictures	Nouns–Action
			Posttest		
High-SES white	.82	2.66	2.52	2.84	4.62
Low-SES black	.72	2.63	1.89	2.57	4.25
			Pretest		
High-SES white	.69	1.08	2.48	3.02	4.02
Low-SES black	.38	.45	2.54	2.45	3.42

Note.—Maximum score per item type = 5.00.

The multivariate tests of most pertinence to the present investigation, of course, are those appropriate for determining whether the effects of the treatment conditions were constant across item types in the two populations. This description, constant effects across item types, is supported by the results of the four relevant multivariate tests; for the high-SES whites and the low-SES blacks, respectively, neither the variance associated with the comparison of practice with control (Fs = 1.50 and 1.61; $df = 4/71$; $p > .05$) nor that associated with training versus the average of practice and control (Fs = .35 and 1.95; $df = 4/71$; $p > .05$) was significant. That is to say, none of the significant contrasts among treatment conditions revealed by analysis of between-subjects variance sources differed as a function of item types. Thus, it is warranted to restate the conclusion that specific training is effective for high-SES white children on the aural posttest and for the low-SES black children on the mixed-list posttest.

DISCUSSION

Four aspects of the present results call for additional comment. The most obvious of these is that elaboration training can, in fact, increase the learning efficiency of high-SES white and low-SES black second-grade children. A second interesting feature of the results is the specificity of the training effect with respect to populations and type of posttest. The content of instruction in the training condition was omnibus in character, involving both visual and verbal skills at both the item and the pair level. Nevertheless, the effects of training were not general. For the high-SES white children, the training effects were discernible only on the more difficult posttest task, the aural list, where neither item nor pair elaboration was provided in the materials.

This result is comprehensible in terms of the notion that the aural list provided a maximum opportunity for the operation of subject-initiated elaborative activity. The problem with this account is that the aural list detected no training effect among the low-SES black children; for this sample, the training condition was superior to the practice and control conditions only on the mixed-list posttest. The one clear inference from this pattern of results is that the effect of training, although positive in both cases, was different for the two populations. It is not possible to discern the specific character of this difference in the present results. Accordingly, the results lend force to the recommendation that in training studies posttest tasks should be selected which can resolve general treatment effects into their components and which elicit a variety of performances sufficient to accomodate the possibility that training may have different effects for different populations.

A third noteworthy aspect of the present results is the populations difference itself. One of the objectives of the training treatment was to remedy deficiencies in the elaborative learning skills of the low-SES black children relative to the high-SES whites. The results indicate that this attempt was, in a practical sense, a failure; that is, on both posttest tasks, the learning efficiency of the low-SES blacks approximated that of the high-SES whites more closely in the control condition than in the training condition. This outcome is not surprising in the case of the aural list since it detected no effect of training for the low-SES black sample but the same pattern is evident on the mixed list which did show a positive effect of training. Thus, implementation of the training treatment used in the present study should be made with full recognition that its likely effect will be beneficial for both populations but more so for high-SES whites than for low-SES blacks.

Finally, still another cautionary note must be sounded about the potential of instruction in elaborative skills for attaining the objective of improving learning efficiency in second-grade children. The data presented in Tables 1 and 2 suggest that the method of presenting learning materials can have a more pronounced effect on learning efficiency in young children than the kind of training experience which they are given; the differences among item types appear larger than the differences among treatment conditions. It may be speculated that for children in the age range sampled, here, the acquisition of designated content can be better assured by careful design of the method of presentation than by attempts to improve the learning skills of the child. It could be that training in elaborative skills would be more productive for older rather than for younger children.

REFERENCES

BOWER, G. H. Mental imagery and associative learning. Paper presented at the Fifth Annual Symposium on Cognition, Carnegie-Mellon University, Pittsburgh, April 1969.

BUGELSKI, B. R. Presentation time, total time, and mediation in paired-associative learning. *Journal of Experimental Psychology,* 1962, **63,** 409–412.

DAVIDSON, R. E., & ADAMS, J. F. Verbal vs. pictorial facilitation of paired-associate learning. *Journal of Experimental Child Psychology,* 1970, **9,** 429–435.

JENSEN, A. R., & ROHWER, W. D., JR. Verbal mediation in paired-associate and serial learning. *Journal of Verbal Learning and Verbal Behavior,* 1963, **1,** 346–352.

JENSEN, A. R., & ROHWER, W. D., JR. Syntactical mediation of serial and paired-associate learning as a function of age. *Child Development,* 1965, **36,** 601–608.

MARTIN, C. J. Associative learning strategies employed by deaf, blind, retarded and normal children. *Educational Research Series,* 1967, **38,** 1–158.

MILGRAM, N. A. Retention of mediation set in paired-associate learning of normal children and retardates. *Journal of Experimental Child Psychology,* 1967, **5,** 341–349. (a)

MILGRAM, N. A. Verbal context versus visual compound in paired-associate learning by children. *Journal of Experimental Child Psychology,* 1967, **5,** 597–603. (b)

MONTAGUE, W. D., & WEARING, A. J. The complexity of natural language mediators and its relation to paired-associate learning. *Psychonomic Science,* 1967, **7,** 135–136.

PAIVIO, A. Mental imagery in associative learning. and memory. *Psychological Review,* 1969, **76,** 241–263.

REESE, H. W. Imagery in paired-associate learning in children. *Journal of Experimental Child Psychology,* 1965, **2,** 290–296.

ROHWER, W. D., JR. Social class differences in the role of linguistic structures in paired-associate learning: Elaboration and learning proficiency. Final report, United States Office of Education Basic Research Project 5-0605, Contract No. OE 6-10-273, November 1967.

ROHWER, W. D., JR., AMMON, M. S., SUZUKI, N., & LEVIN, J. R. Population differences in learning proficiency. *Journal of Educational Psychology,* 1971, **62,** 1–14.

ROHWER, W. D., JR., LYNCH, S., LEVIN, J. R., & SUZUKI, N. Pictorial and verbal factors in the efficient learning of paired associates. *Journal of Educational Psychology,* 1967, **58,** 278–284.

RUNQUIST, W. N., & FARLEY, F. H. The use of mediators in the learning of verbal paired associates. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3,** 280–285.

(Received June 8, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 5, 384–386

TEACHING DISADVANTAGED PRESCHOOL CHILDREN TO THINK CREATIVELY WITH PICTURES[1]

JOE KHATENA[2]
Marshall University

The effects of presenting 118 disadvantaged preschool children with a deliberate training procedure incorporating three creative thinking strategies designed to increase their creative behavior as measured by the figural forms of the *Torrance Tests of Creative Thinking* were explored. The Solomon four-group design was used. Analyses of variance of the posttest scores showed significant main effects for the experimental groups on fluency, flexibility, originality, and elaboration. Significant interaction effects for training and sex, and training and testing, were found only for fluency and elaboration, respectively. It was concluded that preschool disadvantaged children could be taught to think in more creative ways.

Very little has been done by way of experimentation with creative training procedures at the preschool level and nearly nothing substantial has been reported about the effects of providing such nurture to the disadvantaged preschool child despite the considerable efforts of researchers in the field (e.g., Bereiter & Engelmann, 1966; Frost & Hawkes, 1966; Torrance, 1969). Many deliberate training procedures including Osborn's principles of creative problem solving, idea-generating techniques, free association, incubation, breaking away from the obvious and commonplace, transposition, analogy, restructuring, and synthesis have been used to bring about improvement in creative thinking (e.g., Khatena, 1970; Maltzman, Borgartz, & Breger, 1958; Mednick, 1962; Parnes, 1966; Torrance, 1965). However, one experimental study with preschool children using a training program largely rooted to the curriculum was reported by Torrance and associates at Georgia with subjects who could not as a group be considered disadvantaged children (Torrance, Fortson, & Orcutt, 1967). It is the purpose of this study to investigate the effects of teaching disadvantaged preschool children to think in creative ways.

PROCEDURE

Experimental Design

The Solomon four-group design was chosen for the experiment (Campbell & Stanley, 1966). The design required the random assignment of subjects to four treatment groups such that Groups E_1 and C_1 would received the pretest and posttest with E_1 being taught, and Groups E_2 and C_2 would receive the posttest with E_2 being taught.

Subjects

Of the 140 children from seven preschool classes functioning in four different Edgecomb County Schools in North Carolina, 118 children (52 males and 66 females) between the ages of five and six served as subjects. These children were randomly assigned to the four treatment groups. The loss of 22 subjects to problems of placement or their absence while the experiment was in progress reduced the size of the four groups to 27, 30, 31, and 30.

Instruments

The training procedures consisted of teaching children to use three creative thinking strategies, namely, breaking away from the obvious and commonplace, restructuring, and synthesis (J. Khatena, unpublished manuscript, 1969). The subjects were provided with three activities which called for the use of these strategies. The first required them to break away for the perceptual set of the square by making other common objects like the window or door. One possibility that was suggested was a huge bubble with the square as a highlight. For the second and third activities, subjects were provided with a black flannel 6 × 8 inch board and a plastic sac containing 30 pieces (10 of each) shaped as triangles, rectangles and semi-circles. In the second activity the subjects were allowed to use only three triangles, four rectangles, and four semi-circles which had been structured as a man. They were taught to restructure this picture into that of a car with pieces placed on the flannel board. They were then encouraged to use their imagination to restructure the car into other even more interesting and unusual objects, using the same 11 pieces. The third activity allowed the subjects to use all 30 pieces. To illustrate synthesis, a picture of two boys on a see-saw was constructed by the children with the experimenter's help. The subjects were then encouraged to make other pictures which they thought no one else would have thought of by using as many of the 30 pieces as they wished.

---

[1] This study was supported in part by the Georgia Studies of Creative Behavior, University of Georgia, and Marshall University.

[2] Requests for reprints should be sent to Joe Khatena, Department of Educational Foundations, Marshall University, Huntington, West Virginia 25701.

Seven teachers of seven classes, trained by the experimenter to use the instrument, taught their experimental subjects to think in creative ways, using the three creative thinking strategies described. The teachers were themselves trained for 2 hours a day over a period of 2 days prior to the experiment, and the experimental subjects were then trained by these teachers for 1 hour a day over a period of 6 days. A coordinator checked on the uniformity of the training by visiting each of the seven schools during the training period.

Evaluation of the outcome of the training was done by using both of the figural forms of the *Torrance Test of Creative Thinking* (1966). The subjects were measured on four different creative abilities, namely, fluency, flexibility, orginality, and elaboration.

Statistical Treatment

A 2 × 2 × 2 factorial design was used and posttest scores were subjected to analysis of variance to estimate main and interaction effects (Edwards, 1960).

RESULTS AND DISCUSSION

Transfer effects of training are evident in the posttest scores of the experimental subjects on fluency, flexibility, originality and elaboration (Table 1). The difference between the means for fluency of Groups E_1 and C_1 increases from .87 on the pretest to 2.02 on the posttest, while the difference between E_2 and C_2 on the posttest is 4.49, and in both cases in favor of the experimental groups. An analysis of variance of the posttest scores shows significant gains resulting from the training ($F = 4.168$, $df = 1$, $p < .05$), and significant interaction effects between the training and sex variables ($F = 5.289$, $df = 1$, $p < .05$) without significant main effects for sex alone.

For flexibility, the difference of 1.46 between the pretest mean scores for groups E_1 and C_1 increases to 2.21 on the posttest, while E_2 shows an increase of 1.79 above C_2 on the posttest. An analysis of variance of the posttest scores of the four groups shows significant main effects alone for groups E_1 and E_2 ($F = 4.450$, $df = 1$, $p < .05$).

On originality there is hardly any difference between the means of groups E_1 and C_1 on the pretest, but following the training, E_1 almost doubles its mean score on the posttest whereas C_1 shows almost no gain. The difference between the posttest means for E_1 and C_1 was 10.09. As for the posttest mean of group E_2, it is more than double that of C_2 with a mean difference of 9.57. An analysis of variance of the posttest scores shows the main effects of training to be highly significant ($F = 13.511$, $df = 1$, $p < .01$), and although there are some tendencies toward interaction effects between the training and sex variables, these are not significant.

With reference to elaboration, group C_1 shows itself to be somewhat superior on the pretest to E_1 with a difference of .86 between the two means. Following the training, the position is reversed with the posttest mean for group E_1 being higher than the posttest mean for C_1 by 5.37. However, both groups show substantial gains on the posttest, with the gains of group C_1 probably due to the warm-up effects of the pretest. An analysis of variance of the posttest

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR FLUENCY, FLEXIBILITY, ORIGINALITY, AND ELABORATION ON THINKING CREATIVELY WITH PICTURES; FORMS A AND B

Group treatment	Creative ability measure	Pretest		Posttest	
		M	*SD*	*M*	*SD*
E_1 Tested and taught (n = 27)	Fluency	5.07	5.69	10.19	8.37
	Flexibility	4.93	6.10	6.11	5.03
	Originality	10.19	8.55	20.22	14.50
	Elaboration	17.37	17.58	39.00	26.15
C_1 Tested but not taught (n = 30)	Fluency	4.20	3.41	8.17	10.09
	Flexibility	3.47	3.13	3.90	4.03
	Originality	10.10	6.93	10.13	8.34
	Elaboration	18.23	13.42	33.63	28.38
E_2 Taught (n = 31)	Fluency			9.29	6.63
	Flexibility			5.29	3.43
	Originality			18.94	13.19
	Elaboration			41.87	24.95
C_2 Tested (n = 30)	Fluency			4.80	5.83
	Flexibility			3.50	4.21
	Originality			9.37	10.40
	Elaboration			18.07	21.26

mean scores shows significant main effects ($F = 5.375$, $df = 1$, $p < .01$) and interaction effects of training and testing ($F = 4.258$, $df = 1$, $p < .05$), with no significant interaction effects between training and sex.

All in all, the exposure of disadvantaged preschool children to three creative thinking strategies, using nonverbal stimuli proved to be successful in bringing about significant gains in creative behavior as measured by the *Torrance Tests of Creative Thinking* (Figural Forms A and B).

The evidence of this study suggests that disadvantaged children at the preschool level can be taught to think in creative ways not only with experimental but also with everyday learning materials. This cannot be overemphasized. Children between the ages of five and six in some ways can and do think in creative directions with even more freedom than older children. The stage of their cognitive development allows them to handle nonverbal material where it will not permit them to handle the verbal material. What matters is that, while the maturity of their cognitive processes to handle symbolic material has to await development, the children are not denied the opportunity to use the thinking processes themselves. Their propensity is to enter into imaginative experiences more readily at this stage than during the early school years when the cramping and corrosive influences of an expanding intellect and social pressures lead to the first of the several slumps in their creative thinking abilities at the fourth grade level; this factor should be the cue for educators to provide such experiences as will nurture and reward imaginative behavior so that children will have the strength to combat and overcome the negative or neutralizing forces that will inevitably operate upon them in the context of their formal education, a strength that will lead to more effective and permanent learning with the built-in controls for appropriate germination of imaginative productivity.

Insofar as this is applicable to the education of the disadvantaged preschool child, so is it even more penetrating in its implications for the child who is in the context of social and economic advantage. However, as essential as early nurture and reward of creative behavior may seem, it must be seen in the perspective of the continued education of the child if its full significance is not to be lost. It is important that the work of fostering and developing creative ways of mental functioning and behavior should not only be begun with children at the preschool age but also be continued and strengthened through all the levels of formal education.

REFERENCES

Bereiter, C., & Engelmann, S. *Teaching disadvantaged children in the preschool.* Englewood Cliffs: N.J.: Prentice-Hall, 1966.

Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental design for research.* Chicago: Rand McNally, 1966.

Edwards, A. L. *Experimental designs to psychological research.* (rev. ed.) New York: Holt, Rinehart & Winston, 1960.

Frost, J. L., & Hawkes, G. R. (Eds.), *The disadvantaged child.* Boston: Hougton Mifflin, 1966.

Khatena, J. Training college adults to think creatively with words. *Psychological Reports,* 1970, **27,** 279–281.

Maltzman, I., Bogartz, W., & Breger, L. A procedure for increasing word association originality and its transfer effects. *Journal of Experimental Psychology.* 1958, **56,** 392–398.

Mednick, S. S. The associative basis of the creative process. *Psychological Review,* 1962, **69,** 220–232.

Parnes, S. J. *Programming creative behavior.* (Department of Health, Education, and Welfare, United States Office of Education, Title VII, Project N. 500716) Washington, D.C.: 1966.

Torrance, E. P. *Rewarding creative behavior.* Englewood Cliffs, N. J.: Prentice-Hall, 1965.

Torrance, E. P. *Torrance Tests of Creative Thinking: Directions Manual and Scoring Guide.* (Verbal Test Booklets A & B). Princeton, N.J.: Personnel Press, 1966.

Torrance, E. P. Issues in the identification and encouragement of disadvantaged children. *TAG Gifted Children Newsletter,* 1969, **11,** 48–45.

Torrance, E. P., Fortson, L. R., & Orcutt, L. E. The creative-aesthetic approach to school readiness and measured creative growth. Unpublished manuscript, Research & Development Center, University of Georgia, July 1967.

(Received June 12, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 5, 387-394

EFFECTS OF THREE TYPES OF INSERTED QUESTIONS ON LEARNING FROM PROSE[1]

GRAEME H. WATTS[2] AND RICHARD C. ANDERSON[3]
University of Illinois

Three hundred high school seniors answered an inserted question after reading each of five 450-word passages explaining a psychological principle. The question involved identifying an example of the principle or the name of the psychologist associated with the principle. Subjects who received questions that required them to apply the principles to new examples performed significantly better on the posttest than all other subjects, including subjects who received otherwise identical questions that repeated examples described in the text. The group which answered name questions performed worst on the posttest, poorer even than the reading-only control group. It is argued that inserted application questions induced the student to process the text more thoroughly. However, it is not clear whether the facilitative effect is attributable to a "forward," shaping influence or a "backward," review influence.

Recent studies have shown that students learn more from prose passages when they are occasionally required to answer questions inserted within the text (Frase, 1967; Rothkopf, 1966; Rothkopf & Bisbicos, 1967). When questions are repeated on the posttest people who answered these questions during the passage perform substantially better than people who only read the passage. People who answer questions following (but not before) sections of text also perform somewhat better on new posttest questions. This indirect effect has been attributed to the control which the questions exercise over "mathemagenic activities," that is, the processing the learner gives to the material.

As interesting as the research has been, there is still little in the literature on questioning that is of practical value to education. The problem lies in the kind of questions used in studies reported to date. Typically, the questions and the answers have been lifted verbatim from the instructional materials and usually have involved names, dates, technical terms, or quantities. Most important, educational goals do *not* entail verbatim recall. Whether or not a student is able to accurately recognize or repeat strings of words is educationally trivial. It remains to be seen whether the results of research with questions which depend upon understanding of a concept or principle will parallel the findings obtained with questions involving literal repetition of instructional statements.

There is reason to believe that questions which force the student to go beyond the literal content of instruction will facilitate learning. The idea is that questions that demand more than verbatim recall will promote deeper processing of the instructional materials. Recent studies suggest that procedures which force subjects to comprehend the meaning of sentences facilitate sentence learning (Anderson, Royer, Kulhavy, Thornburg, & Klemt, 1971; Bobrow & Bower, 1969). It is reasonable to suppose that questions that require students to apply the concepts or principles described in a passage can have similar effects.

The main purpose of this study was to investigate the effects of questions that require students to apply what they have read to some new situation. For each of five passages explaining a psychological principle, a set of questions was prepared according to a single recipe: Given the name of a principle, the student had to select an example of the principle from among a list of four possible examples. In half the questions the correct alternative repeated an example from the text. The remaining questions entailed a new example of the principle, different from any described in the text. A third type of question, comparable to the questions employed in previous research, required the student to select the name of the psychologist associated with a principle.

---

[1] The research reported in this paper was supported in part by the Advanced Research Projects Agency through the Office of Naval Research under Contract ONR Nonr 3985(08).

[2] Now at Balmain Teachers College, Rozelle, New South Wales, Australia.

[3] Requests for reprints should be sent to Richard C. Anderson, Department of Psychology, University of Illinois, Urbana, Illinois 61801.

METHOD

Subjects

The subjects were 300 high school seniors. Students from classes designated as below average in achievement level were not used and no pupil who had taken a psychology class was included in the study.

Materials

Passages. Five passages were compiled which described the principles of classical conditioning, extroversion-introversion, intermittent reinforcement, displacement, and drive reduction.[4] The passages were very similar in organization and style. The first paragraph of each passage presented a nontechnical account of some situation illustrating the principle. In the second paragraph appeared the name of the psychologist commonly associated with the principle, a general explanation of the principle, and a second illustration. The third paragraph introduced an additional technical term related to the principle and a concluding summary sentence. The passages were about 450 words in length.

Questions. Five questions were prepared for each passage. All were written in a multiple-choice format and conformed to one of the following descriptions. In repeated-example (RE) questions, the correct multiple-choice alternative exactly reproduced a situation described in the text to illustrate a principle. Question RE_1 repeated the example that appeared in the opening paragraph of the passage whereas questeion RE_2 repeated the example in the second paragraph. The two application (A) questions for each passage were identical to the repeated-example questions except that the correct alternative in application questions described an example of the concept or principle different from any example employed in the passage. To make concrete the distinction between repeated-example and application questions, an example of displacement which appeared in the text and a repeated-example question was, "John failed to make the basketball squad because he was too small so he practiced hard to win a cheerleader's position." The example of displacement contained in one of the application question was, "When punished for sucking his thumb Timmy stopped that habit but spent hours chewing gum given him by his grandmother."

Each of the three distractors in a repeated-example or application question described an example of a certain incorrect interpretation of the concept or principle. The same three incorrect interpretations were used to generate distractors for all the repeated-example and application questions for a passage. Thus, the distractors used in any set of repeated-example and application questions could be freely exchanged without affecting the relationship between the distractors and the correct alternative.

In name (N) questions, the correct alternative was the name of the psychologist identified with the principle in the first sentence of the second paragraph of each passage. Only one psychologist was mentioned in any passage. The name questions contained five response alternatives. Three of the distractors were names introduced in other passages, and one was the name of a psychologist not mentioned during instruction.

Design

The design involved one between- and three within-subject factors. The between-subject factor, hereafter called treatment, was the type of question inserted after each passage. Group RE_1 was required to answer the RE_1 question prepared for each passage, Group RE_2 the RE_2 question, Group A_1 the A_1 question, and so on. A reading-only control group (Group C) simply read through the passages and took the posttest.

The three within-subject factors were subject matter (five principles), position of a given passage in the instructional sequence (one to five), and question type on the posttest (five types). The sequence of presentation of passages was ordered in accordance with a Latin square with each arrangement presented equally in every treatment group.

Since the storage requirements for the data analysis far exceeded the limits of computer programs available for analysis of variance and since there was no interest in any possible interaction between the subject matter and position factors, the main analysis of variance for posttest scores was prepared by combining two separate analyses: (*a*) involving treatment, position, and question type, and (*b*) involving treatment, subject matter, and question type. The total analysis, therefore, contained all terms except those involving Subject Matter × Position interactions. In the two separate analyses there were no differences among the within-subjects error terms so these were combined to form the pooled error in the total analysis. This residual included all terms containing a Subject Matter × Position interaction.

Since a large number of multiple comparisons was anticipated, control of Type 1 error rate required some attention. Following Petrinovich and Hardyck (1969), the Tukey B procedure was used because it fixes experimentwise error rates at conventional levels while being somewhat more sensitive to real differences between groups than the extremely conservative Scheffé procedure.

---

[4] For copies of the passages and questions, order NAPS Document No. 01560 from ASIS National Auxiliary Publications Service, c/o CCM Information Sciences, Inc., 22 West 34th Street, New York, New York 10001; remitting $2.00 for microfiche or $5.50 for photocopies.

Procedure

The study was run in the school auditorium in five shifts of approximately 60 subjects. There were approximately equal numbers of subjects from each treatment group in each shift. The five passages were mimeographed and stapled into booklets. The subjects were randomly assigned to treatment groups simply by distributing the booklets which previously had been arranged in random order. The instructions stressed that the materials were to be read once only, without taking notes or marking the booklets, and that the subjects should be prepared to take a test when finished. No time limit was imposed but the subjects recorded the time showing at the front of the room whenever they came to blue time sheets inserted throughout the booklets. The inserted question was answered in the booklet after each passage. The subject was not permitted to examine the passage while considering the question.

The posttest began as soon as the passages were completed. It contained 25 questions, the five questions prepared for each of the five passages. There were five different random orders of test items assigned equally often at random to subjects within treatment groups.

Results

Posttest Performance

Table 1 summarizes the analysis of variance for all experimental groups on the 25-item posttest. All main effects were significant beyond the .01 level.

Mean posttest percentages under the six treatments are presented in Table 2. Both groups answering application questions during instruction performed significantly ($\alpha = .05$) better than all other groups. The most marked differences were between the two Application groups and the RE_1 and Name groups.

Figure 1 shows the position means, indicating the classical serial position effect. A trend analysis revealed a significant quadratic component ($F = 23.40$, $df = 1/6936$, $p < .01$) while no other component was significant (all Fs < 1.0).

As expected, there was considerable variation in performance on the different types of questions (see Table 2). While all differences proved significant ($\alpha = .01$), obviously the repeated-example questions were the easiest to answer correctly. Mean performance was about the same for four topics (displacement = 60.9%, drive reduction = 64.3%, extroversion–introversion = 62.2%, intermittent reinforcement = 62.0%), while performance on questions relating to the classical conditioning topic (80.0%) diverged significantly ($\alpha = .01$).

TABLE 1

Analysis of Posttest Variance

Source	*df*	*MS*	*F*
Between subjects			
Treatment (T)	5	3.51	5.93*
Subjects within groups	294	.59	
Within subjects			
Position of Instructional Passage (P)	4	1.03	6.12*
T × P	20	.19	1.14
Question Type (Q)	4	44.65	264.54*
T × Q	20	.81	4.79*
Subject Matter (SM)	4	9.58	56.76*
T × SM	20	.19	1.16
P × Q	16	.32	1.91
T × P × Q	80	.16	.95
SM × Q	16	2.86	16.93*
T × SM × Q	80	.18	1.08
Pooled error	6936	.17	

* $p < .01$.

The significant Treatment × Question Type interaction ($p < .01$) merits more detailed inspection. Mean performance of each treatment group is summarized for each question type in Table 2. With respect

TABLE 2

Mean Percentage Correct for Each Treatment on the Five Question Types Answered on the Posttest

Question type	Treatment						All treatments
	RE_1	RE_2	A_1	A_2	N	C	
RE_1	88.0	84.4	88.4	90.0	79.6	88.0	86.4
RE_2	75.6	84.0	85.6	84.8	79.2	81.6	81.8
A_1	48.0	50.4	70.4	63.6	42.8	48.0	53.9
A_2	58.8	57.6	70.4	71.6	47.6	52.8	59.7
N	36.4	46.8	48.8	50.4	52.8	50.4	47.6
All types	61.4	64.6	72.7	72.1	60.4	64.2	65.9

Note.—Abbreviations: RE = repeated example, A = application, N = name, C = control.

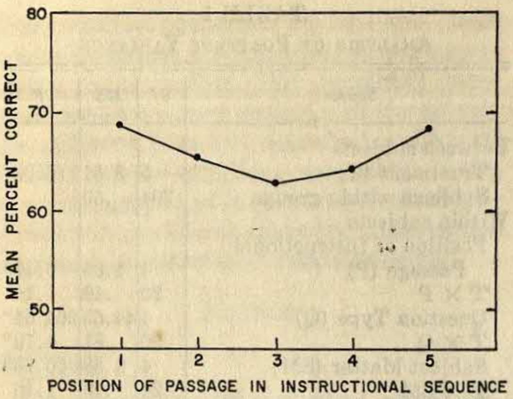


FIG. 1. Performance on the posttest as a function of serial position of passage during instruction.

to A_1 questions both the A_1 and A_2 treatment groups enjoyed a significant ($\alpha = .01$) advantage over all other treatment groups. Differences between the remaining four groups were not significant. The A_1 treatment group had been asked to answer the A_1 question immediately following each passage; therefore it is not surprising that this group performed well on A_1 posttest questions. What is interesting is the good performance of the A_2 group on A_1 questions. Likewise, the A_1 group showed positive transfer on the A_2 questions. While there was virtually no difference between the A_1 and A_2 treatment groups on A_2 questions, they both displayed a significant ($\alpha = .01$) advantage over the remaining groups.

Performance on the name questions was relatively poor. The RE_1 group fell significantly ($\alpha = .05$) below all other groups but no other differences approached this significance level. The fact that groups required to answer repeated-example and application questions during instruction did not perform very well on name questions was not unexpected, but it is puzzling that the Name group did not prove superior on name questions. Each subject in the Name group had seen all the name questions before and had been remarkably successful in answering those questions during instruction. On the face of it they should have outperformed the other groups on this question type on the posttest.

There were no differences which prove significant with respect to repeated-example questions. Apparently most subjects had little difficulty in recognizing at least some feature of the examples used in the passages and so were able to select the correct alternative on these questions. On such relatively easy questions there was evidently no advantage accruing to the RE_1 and RE_2 groups from having answered the repeated-example questions during instruction.

The significant Subject Matter × Question Type interaction is uninteresting and will not be discussed in detail. For all subject matters, the repeated-example questions were answered the most successfully, but the name and application questions did not always fall in the same rank order.

Performance on Inserted Questions

An analysis of performance on inserted questions showed a significant main effect for treatment ($F = 31.81$, $df = 4/245$, $p < .01$) but not for position of passage ($F = 1.78$, $df = 4/980$) nor Treatment × Position ($F = 1.43$, $df = 16/980$). The subjects in the Name group showed nearly perfect performance (99.6%) over the five questions, followed by high scores from the RE_1 (90.4%) and RE_2 (87.2%) groups, with the application question groups trailing off sharply ($A_2 = 73.2\%$, $A_1 = 62.0\%$). Post-hoc comparisons showed that all differences except between groups RE_1 and RE_2 were significant ($\alpha = .01$). It is worth noting that this rank ordering of groups with respect to success on inserted questions during instruction exactly reverses the ranking of the groups on the posttest (see Table 2). Obviously, success in answering questions during instruction does not in itself guarantee a similar level of success on a posttest.

Time on Inserted Questions

Differences in the time spent on answering inserted questions were significant ($F = 94.12$, $df = 4/245$, $p < .01$). There was a direct correspondence between difficulty level of inserted questions and time spent on these questions, with name questions given the shortest mean time (11 seconds), then the repeated-example questions ($RE_1 = 36$ seconds, $RE_2 = 37$ seconds), and finally the application questions were given the longest time ($A_1 = 47$ seconds, $A_2 = 47$ seconds). All groups reduced time on questions over the instructional period ($F = 68.43$, $df = 4/980$, $p < .01$) but there was no significant Treatment Group × Position of Passage interaction ($F = 1.38$, $df = 16/980$).

Time on Passages

Although there was a significant decrease in time spent on each passage over the instructional sequence ($F = 17.72$, $df = 4/1176$, $p < .01$) there were no significant differences between groups in the mean total time spent on all passages ($F = 1.22$, $df = 5/294$). The Treatment Group × Position of Passage interaction was significant ($F = 1.93$, $df = 20/1176$, $p < .01$). These means are presented in Figure 2. A trend analysis showed significant linear ($F = 5.81$, $df = 1/1176$, $p < .01$) and cubic components ($F = 4.20$, $df = 1/1176$, $p < .01$). From Figure 2 it is clear that it is the Name group means which diverge from the means of the remaining groups. While all groups gave much the same time to the first two passages, the Name group more rapidly decreased time spent on the final three passages.

Time on Posttest

There were significant differences between groups in time spent on the posttest ($F = 7.28$, $df = 5/294$, $p < .01$). Group C (13.2 minutes) and Group N (13.1 minutes) took the longest mean times, followed by the two repeated-example groups (Group $RE_1 = 11.8$ minutes, Group $RE_2 = 11.3$ minutes), while the two application groups (Group $A_2 = 11.1$ minutes, Group $A_1 = 10.8$ minutes) took the shortest time on the posttest. It is worth noting that this rank ordering reverses the rank ordering of groups on mean times given to inserted questions during instruction. On the posttest the Name group, and to a lesser extent the RE_1 and RE_2 groups, apparently had to devote more time to selecting an answer for the more difficult application questions for which they were not as well prepared as the A_1 and A_2 groups.

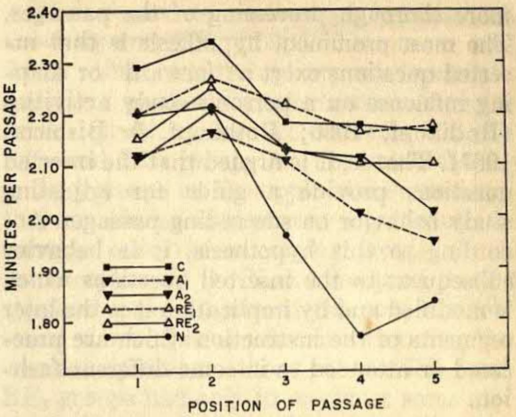


FIG. 2. Mean times taken by each treatment group on the five instructional passages.

DISCUSSION

Performance on a posttest was studied as a function of the type of question inserted into the instructional sequence. Students who answered application questions during instruction demonstrated the best overall performance. They were at no disadvantage on the repeated-example and name questions while at the same time demonstrating a striking superiority on both the repeated and alternate application questions.

It is interesting to note that the Application groups had the least success of any of the groups in answering the inserted questions while studying the passages. This fact leaves open the possibility that the apparent advantage from answering application questions during instruction was not related to the type of question asked, but simply to an increase in motivation due to having to answer more difficult questions. Perhaps difficult questions of *any* type would have had the same effect, though available evidence seems to indicate that it is easy questions which have long-term facilitative effects (Marso, 1969). By manipulating the distractors, it should be possible to control the difficulty level of the application and repeated-example questions and put this hypothesis to the direct test.

But taking the data as they now stand it seems reasonable to suggest that the subjects answering application questions during instruction were forced to engage in more thorough processing of the passages. The most prominent hypothesis is that inserted questions exert a "forward" or shaping influence on a person's study activities (Rothkopf, 1966; Rothkopf & Bisbicos, 1967). That is, it is argued that the inserted questions provide a guide for adjusting study behavior on succeeding passages. According to this hypothesis, it is behavior subsequent to the inserted questions which is modified and by implication it is the later segments of the instruction which are processed or attended to in some different fashion.

Auxiliary data from the present experiment bear upon the shaping hypothesis. If answering questions does affect study behavior on subsequent passages, then this effect should be more evident on posttest questions related to passages appearing near the end of the instructional sequence, when the type of question will be most predictable for the subjects. However, consider the following posttest means (ordered according to passage position) for each treatment group on the same questions they answered during instruction: RE_1 percentage = 88, 92, 84, 86, 90; RE_2 percentage = 88, 78, 88, 82, 84; N percentage = 66, 40, 44, 40, 74; A_1 percentage = 70, 66, 58, 74, 84; A_2 percentage = 74, 74, 72, 66, 72. In no group was there any suggestion of increased success over passages, but rather in nearly every case the classical serial position effect appeared. That this position effect holds when performance is summed over all posttest scores (refer to Figure 1) provides an interesting parallel to basic studies in verbal learning.

A more direct effect of the forward influence of questions should be evident in a subject's success in answering the inserted questions over passages. The RE_1, RE_2, and Name groups were close to the ceiling on the first passage, but there was an opportunity for the A_1 and A_2 groups to show an upward trend over inserted questions. The mean percentages for the A_1 group (percentage = 66, 56, 46, 66, 76) did show some upward swing over the last two passages but this was not true for the A_2 group (percentage = 74, 72, 76, 72, 72). Thus, performance over passages on inserted questions was not consistent with the hypothesis that questions have a forward influence.

The time data are also relevant to the forward influence hypothesis. All groups reduced the time spent on each passage as they worked through the booklet, but there was no significant difference between the groups on the total time given to all passages. The Name group diverged sharply from the others and this suggests that the subjects in this group may have altered their strategy as a consequence of receiving easy questions. However, this argument cannot be extended to other groups. The A_1 and A_2 groups, which had the most difficult questions, actually took less time over passages than did the RE_1 and RE_2 groups.

The failure to find any real support for the forward influence hypothesis suggests consideration of an alternative hypothesis, that of a "backward" or review influence. In this study, the subjects who were required to answer repeated-example or name questions could have done so correctly by entering predictable information into short-term storage until the end of the passage where they simply checked off the answer they were expecting to have to give. For these subjects there would have been little need to review the content of the passage before answering the inserted question. However, subjects with application questions probably had to consider each alternative as a posssible example of the principle and in so doing would have had to mentally review the content of the passage before making a decision. Furthermore, those who received application questions undoubtedly had to bring to mind the meaning of key words in passages, either as they read or while they considered questions, if they were to correctly identify new examples. The name and repeated-example questions could be answered by "rote."

The fact that subjects in both the A_1 and A_2 groups took longer to answer inserted questions offers some indirect support for the hypothesis that the application questions caused them to mentally review the passages. However, it could be that the application questions simply took longer to read. The subjects who received name questions had only five names to read. The RE_1

and RE_2 groups could stop reading just as soon as they came to the alternative which matched a statement in the passage. This may explain why these groups took so little time on their questions. On the other hand, the A_1 and A_2 groups would have been more likely to have at least read all the alternatives.

The answer to whether the direction of the influence from inserted application questions is primarily forward or backward will have to wait further study. While the present data do not disprove the forward influence hypothesis, they do not offer much support. Rather, they leave open the distinct possibility, and one which the authors find attractive, that what is important is the character of the post-passage review and processing that a question induces.

While the posttest performance of the A_1 and A_2 groups was as good as, or better than, any other group on *every* question type, the position was almost reversed for the Name group. It is particularly surprising in the light of many studies reporting a posttest advantage on questions practiced during instruction (e.g., Rothkopf, 1966) that the Name group failed to perform any better on name questions than the other groups. That they did give special attention to the name of the psychologist mentioned in each passage is evidenced by their almost perfect record on inserted questions despite a significant decrease in time given to succeeding passages. However, if it can be argued that the subjects only give as much effort as the task demands (Anderson, 1970; Underwood, 1963), then the subjects in the Name group could have successfully answered the inserted questions by doing no more than holding the name of the psychologist in short-term storage until the end of the passage where that name was checked off. They were never required to link a psychologist's name with the stimulus term (name of the principle). Consequently, even on a multiple-choice posttest question, subjects were unable to recognize the correct alternative which they had previously selected because there had never been any requirement to link stimulus and response terms, but only to hold the response term in short-term storage.

With regard to posttest performance of the RE_1 and RE_2 groups, there is no evidence that answering repeated-example questions promoted performance on application questions. In fact, not only did the RE_1 and RE_2 groups fail to match the A_1 and A_2 groups on the application questions, they also failed to show any significant advantage over the remaining groups (with the exception of the Name group on the A_2 question type). Again, if the subjects followed a line of least effort in answering the inserted questions, those in the RE_1 and RE_2 groups had only to recognize some element of a situation described in the passage they had just read. Certainly there was no need for subjects in the Repeated-Example groups to review the content of the passage or to transform that content to fit a new situation. Consequently, it is argued, these subjects were at a disadvantage on application questions where they could no longer rely on recognition of a situation previously described—and this despite the fact that the questions were constructed in exactly the same form with only the details of the example in the correct alternative changed.

The main purpose of this study was to investigate the effect of questions which require students to apply what they have studied to some new situation. It is quite clear that the insertion of application questions not only had a beneficial practice effect when the same questions were repeated on the posttest, but also led to superior performance on new application questions. The authors believe that answering application questions facilitates later performance by encouraging students to process the content of the instruction more thoroughly, in fact to transform it, in the effort to apply it in a new situation.

REFERENCES

ANDERSON, R. C. Control of student mediating processes during verbal learning and instruction. *Review of Educational Research*, 1970, **40,** 349–369.

ANDERSON, R. C., ROYER, J. M., KULHAVY, R. W., THORNBURG, S. D., & KLEMT, L. L. Thematic prompting in paired associate learning. *Journal of Educational Psychology*, 1971, **62,** 315–321.

BOBROW, S. A., & BOWER, G. H. Comprehension and recall of sentences. *Journal of Experimental Psychology*, 1969, **80,** 455–461.

FRASE, L. T. Learning from prose material: Length of passage, knowledge of results, and position of questions. *Journal of Educational Psychology*, 1967, **58**, 266–272.

MARSO, R. N. The influence of test difficulty upon study efforts and achievement. *American Educational Research Journal*, 1969, **6**, 621–632.

PETRINOVICH, L. F., & HARDYCK, C. D. Error rates for multiple comparison methods. *Psychological Bulletin*, 1969, **71**, 43–54.

ROTHKOPF, E. Z. Learning from written material: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal*, 1966, **3**, 241–250.

ROTHKOPF, E. Z., & BISBICOS, E. E. Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology*, 1967, **58**, 56–61.

UNDERWOOD, B. J. Stimulus selection in verbal learning. In C. N. Cofer and B. S. Musgrave (Eds.), *Verbal behavior and learning*. New York: McGraw-Hill, 1963.

(Received June 17, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 5, 395-399

MEANINGFUL PROCESSING OF SENTENCES[1]

RICHARD C. ANDERSON,[2] SHEILA R. GOLDBERG, AND JANET L. HIDDE
University of Illinois

In two experiments involving 91 college undergraduates, subjects who filled blanks at the ends of sentences they read aloud learned significantly more than subjects who read whole sentences. The explanation for this result is that to fill a blank requires a person to comprehend the other words in a sentence whereas a person can "read"—that is, render into speech—a complete sentence without comprehending the words. The second experiment ruled out the alternative explanation that the facilitation associated with filling blanks was due to specific transfer.

There is reason to believe that there are several stages or levels of processing that must occur if a person is to learn from written verbal materials. Two of these stages are auditory encoding and semantic encoding (Anderson, 1970). Auditory encoding consists of rendering words into implicit (or explicit) speech. Semantic encoding is the bringing to mind of meaningful representations of words. It is impossible at the present time to be precise about the nature of semantic encoding, but there is no doubt that it often entails the "mental image" of the things and events named by the words (cf. Bower, in press; Paivio, 1969).

Evidence for auditory and semantic encoding stages comes from research on memory. Most errors in short-term memory arise from confusions between sounds (cf. Wickelgren, 1965, 1966) whereas errors in short-term memory due to confusions in meaning are relatively rare (Baddeley, 1964). In long-term memory, on the other hand, semantic confusion is a more important source of interference than acoustical confusion (Baddeley, 1966; Baddeley & Dale, 1966).

Semantic encoding does not inevitably follow auditory encoding. In simple terms, it is argued that people often "read" without bringing to mind the meaning of the words they are speaking. Our working assumption is that people tend to do no more processing than the task requires (Anderson, 1970). If this analysis is correct, procedures that induce the reader to comprehend words rather than merely say them will facilitate learning.

Bobrow and Bower (1969) have completed several experiments which suggest that procedures which cause subjects to comprehend the meaning of sentences strongly facilitate learning. In one experiment, the subject was instructed to compose a sentence which was a sensible continuation of each sentence he saw. For example, if he saw the sentence "The farmer discovered a diamond," a sensible continuing sentence might be "He sold it to a jeweler and used the money to buy a tractor." Despite the fact that they were vulnerable to interference from the sentences they had created themselves, the subjects in the continuation condition recalled twice as many predicate nouns, given the subject nouns as retrieval cues, than control subjects required to read each sentence aloud three times.

Under some conditions, questions to be answered or blanks to be completed, as in a self-instructional program, may get the reader to process with understanding the material which accompanies the question or blank. Consider the incomplete statement, "Elevators stop at every ———." To complete the sentence with the word "floor" requires a person to bring to mind, in however fleeting a form, a meaningful representation of the rest of the sentence. Simply translating the printed words into speech will not suffice, because the mere sound of the other words can not evoke floor. Floor is semantically rather than acoustically related to the rest of the sentence.

---

[1] The research reported in this paper was supported in part by the Advanced Research Projects Agency through the Office of Naval Research under Contract ONR Nonr 3985(08).

[2] Requests for reprints should be sent to R. C. Anderson, Training Research Laboratory, 226 Education Building, University of Illinois, Urbana, Illinois 61801.

Anderson, Royer, Kulhavy, Thornburg, and Klemt (1971) obtained some indirect evidence that blanks in sentences facilitate learning. In their first experiment, paired associates were learned as readily from simple, unprompted sentences (e.g., "SIG are yellow") and unadorned pairs (e.g., "SIG yellow") as from "thematically prompted" sentences (e.g., "Before turning red, traffic SIG are yellow") displayed as a whole. In two subsequent experiments, the prompted sentences were initially exposed with blanks in place of the response terms in order to induce complete processing of the sentences. Thematic prompts enhanced learning under these conditions. However, this result did not prove that the blank in place of the response term was a critical part of the procedure since none of the experiments directly compared learning from otherwise identical sentences with and without blanks. Herein lies the purpose of the present studies.

EXPERIMENT I

Method

Subjects. The subjects were 43 undergraduate women who participated to fulfill a requirement in an introductory educational psychology course. They were randomly assigned to conditions upon appearance for the experiment.

Sentences. Fifty-five sentences were prepared such that the last word of each sentence seemed to be determined by the other words in the sentence. These sentences, with a blank in place of the final word in each sentence, were presented on mimeographed sheets to 75 subjects following their participation in a free recall experiment. They were instructed to fill each blank with the word that most obviously fit the meaning of the sentence. Selected for use in the experiment were 24 sentences which 99% to 100% of the subjects in the norming group completed with the words intended by experimenters. Figure 1 contains several examples of the sentences.

Procedure. The Blank group received each of 24 sentences with a blank in place of the last word. The sentences were displayed as a whole for the No Blank group. The subjects in both groups read each sentence aloud during the 4-second presentation interval. Those in the Blank group were asked to supply the word which most obviously fit the blank as they read.

Study and test trials alternated until three of each had been completed. Within both types of trials, items were presented at a 4-second rate on a Lafayette Model 303A memory drum. On test trials each subject noun was presented as a retrieval cue; the subject responded orally with the last word in the sentence, which, of course, subjects in the Blank group had never seen. There were three randomizations of the sentences and three randomizations of the subject nouns. One third of the subjects began with each randomization.

Results

A strict scoring system was employed in which only words which completed sentences in the No Blank condition were counted correct. Also tried was a lenient scoring procedure in which a word was counted correct if it matched the word given during the preceding study trial. The lenient scores of the Blank group were 1% higher per trial than the strict scores.

Figure 2 shows mean percent recall (strict scoring). The analysis of variance indicated no significant effect for treatment ($F = 3.80$, $df = 1/41$, $p > .05$). However, there were significant effects due to trials ($F = 313.32$, $df = 2/82$, $p < .01$) and the Treatment × Trials interaction ($F = 12.10$, $df = 2/82$, $p < .01$). Trial by trial comparisons indicated that the Blank and No Blank groups differed significantly on the

A circle has neither a beginning nor ________.

Diamonds are a girl's best ________.

Mother bakes delicious apple ________.

The astronauts were received at the White ________.

Dictionaries list words in alphabetical ________.

The physician noted the time on his wrist ________.

FIG. 1. Examples of incomplete sentences in which the last word is determined by the rest of the sentence.

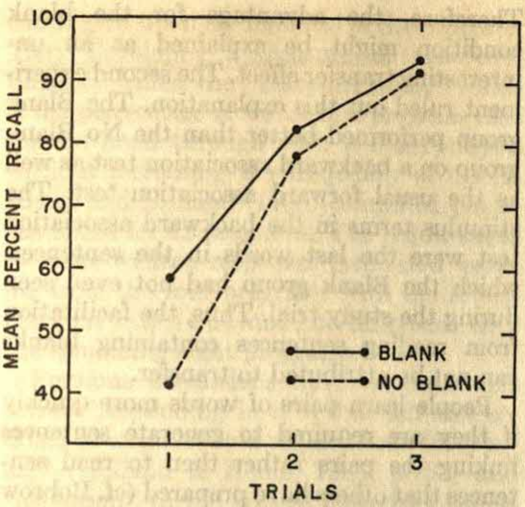


FIG. 2. Mean percentage of recall in Experiment I.

first trial ($t = 3.61$, $df = 41$, $p < .01$), but not on the second or the third.

EXPERIMENT II

In Experiment I, a group that read sentences with blanks in place of the response terms outperformed, at least on the first trial, a group which read entire sentences. The argument is that a blank forces semantic encoding of the other words in a sentence, and that semantic encoding is the precursor to learning. There is, however, at least one alternative interpretation. It could be that the advantage of the Blank group was due to the fact that there was a greater similarity between conditions of learning and conditions of testing for the Blank group than the No Blank group. During learning, the subjects in the Blank group were required to *produce* the final word of each sentence, while subjects in the No Blank group *read* the final word. On the test, both groups were required to *produce* the final word. Thus, the apparent facilitation from requiring blanks to be filled may have been nothing more than transfer. Experiment II investigated the transfer explanation. A backward association test was included in which the subject was required to produce the subject noun of each sentence given the last word of the sentence as a retrieval cue. During study both groups *read* the subject nouns. Furthermore, since subjects in the Blank group did not see the last word of any sentence during study, on the backward test there was greater dissimilarity between the training and test stimuli in the Blank than the No Blank group. Hence, an advantage accruing to the Blank group could not be attributed to specific transfer. Experiment II was also designed to provide a more sensitive test of the effects of the Blank condition. By the second trial of Experiment I the difference between the Blank and No Blank groups was negligible, apparently because of a performance ceiling. In Experiment II there was one study trial on each of three lists, instead of three trials on one list, in order to raise the ceiling.

Method

Subjects. The subjects were 48 paid undergraduate volunteers, randomly assigned to conditions upon appearance for the experiment.

Sentences. A new group of 30 subjects normed 99 additional sentences following participation in a free-recall experiment. Three lists of 23 sentences were constructed containing sentences which 97%–100% of the subjects in one or the other of the norming groups completed with the words intended by the experimenters.

Design and procedure. Half the subjects received the Blank treatment and half received the No Blank treatment. As in the first experiment subjects in both groups read each sentence aloud as it appeared, with those in the Blank group giving the word that went in the blank. All subjects received three lists; one third began with each list. There was one study trial on each list at a 4-second rate, followed by two test trials, also at a 4-second rate. On the forward test trial, each subject noun was presented and the last word of the sentence was required as the response. On the backward test trial, the last word of the sentence was presented and the subject noun was required as the response. Half the subjects received the forward test before the backward test. The order was reversed for the remainder. In all other respects the procedure was the same as in Experiment I.

Results

Reported here are analyses of strictly scored recall in which only words that completed sentences in the No Blank condition were counted correct. As in Experiment I, the lenient scores of the Blank group averaged 1% higher than the strict scores.

The Blank group (73.8%) performed significantly better than the No Blank group (61.7%), ($F = 10.60$, $df = 1/44$, $p < .01$). Treatment entered into two significant inter-

TABLE 1

MEAN PERCENTAGE OF RECALL IN EXPERIMENT II

Group	First test		Second test	
	Forward	Backward	Forward	Backward
Blank	70.8	68.7	82.2	73.6
No Blank	64.7	47.6	62.9	71.7

actions. The first was Treatment × Test Sequence ($F = 4.80$, $df = 1/44$, $p < .05$) which appeared because the Blank group showed a greater overall advantage over the No Blank group when the backward test was given first than when the forward test was given first. The Treatment × Type of Test × Test Order interaction was significant ($F = 4.96$, $df = 1/44$, $p < .05$). As can be seen in Table 1, the Blank group had the greatest relative advantage on the backward test when it was given first but on the forward test when it was given second.

Performance was higher on the forward test (70.2%) than on the backward test (65.4%), ($F = 50.80$, $df = 1/44$, $p < .01$) and on the second test (72.6%) following a list rather than the first (63.0%), ($F = 203.83$, $df = 1/44$, $p < .01$). Otherwise, the only significant terms were List (*not* List Order, which was not significant itself and did not enter into any significant interactions), List × Test Sequence, and List × Type of Test. These effects will not be discussed since they are of no interest.

DISCUSSION

The purpose of the present research was to show that a procedure that forces meaningful processing of sentences facilitates learning. In both experiments people who filled blanks as they read learned more than people who read whole sentences. The explanation for this result is that to fill a blank requires semantic encoding of the other words in a sentence whereas a person can "read"—that is, render into speech—a complete sentence without semantic encoding.

A possible alternative explanation is that there was greater similarity between the conditions of learning and testing under the Blank than under the No Blank condition. Therefore, the advantage for the blank condition might be explained as an uninteresting transfer effect. The second experiment ruled out this explanation. The Blank group performed better than the No Blank group on a backward association test as well as the usual forward association test. The stimulus terms in the backward association test were the last words in the sentences, which the Blank group had not even seen during the study trial. Thus, the facilitation from reading sentences containing blanks can not be attributed to transfer.

People learn pairs of words more quickly if they are required to generate sentences linking the pairs rather then to read sentences that others have prepared (cf. Bobrow & Bower, 1969). This could be interpreted to mean that as they generate sentences people select or create mediators that have idiosyncratically high associative strength. However, the fact that reading incomplete sentences enhanced learning in the present studies cannot be interpreted in terms of idiosyncratic associative strength (idosyncratic vividness, idiosyncratic elaboration, etc.) because everyone completed the sentences in the same way. The implication is that the important aspect of constructing a sentence is not so much the personalized sentence that results as it is the process of giving meaningful representation to the words. The further implication is that people can learn as readily from sentences written by others as they can from ones they have had a part in constructing, provided techniques can be found to get them to process the sentences fully.

A sentence missing a word highly determined by the rest of the sentence bears a strong resemblance to a frame from a self-instructional program. In the light of the results of the present research, it is surprising, therefore, that many studies have found that students who write answers to frames in a program learn no more than students who "think" the answers or read the frames with the blanks filled in (e.g., Alter & Silverman, 1962; Stolurow & Walker, 1962; Tobias & Weiner, 1963). However, Kemp and Holland (1966) were able to show that the requirement to write answers is facilitative if, and only if, correct answers are contingent upon the material in the program. They developed a measure of this contingency called the "blackout ratio," which is the percentage of words in a program that can be lined through with a black crayon without increasing the frequency of errors. Kemp and Holland (1966) computed blackout ratios for 12 programs. The requirement to make written responses facilitated learning from the four programs with the lowest blackout ratios but made no difference with the remaining eight programs.

Previous discussions have stressed that correct responding in a verbal self-instructional program should require "careful reading" (Holland & Kemp, 1968, p. 92) and "attention" to the critical material (Anderson, Faust, & Roderick, 1968). The present research gives additional perspective: Carefulness and attention probably are not enough; it seems likely that if the requirement to fill blanks in a program is to be beneficial, a correct response must be contingent upon *understanding* the critical material.

REFERENCES

ALTER, M., & SILVERMAN, R. E. The response in programed instruction. *Journal of Programed Instruction*, 1962, **1**, 55–78.

ANDERSON, R. C. Control of student mediating processes during verbal learning and instruction. *Review of Educational Research*, 1970, **40**, 349–369.

ANDERSON, R. C., FAUST, G. W., & RODERICK, M. C. "Overprompting" in programmed instruction. *Journal of Educational Psychology*, 1968, **59**, 88–93.

ANDERSON, R. C., ROYER, J. M., KULHAVY, R. W., THORNBURG, S. D., & KLEMT, L. L. Thematic prompting in paired-associate learning. *Journal of Educational Psychology*, 1971, **62**, 315–321.

BADDELEY, A. D. Semantic and acoustic similarity in short-term memory. *Nature*, 1964, **204**, 1116–1117.

BADDELEY, A. D. The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 302–309.

BADDELEY, A. D., & DALE, H. C. A. The effect of semantic similarity on retroactive interference in long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 417–420.

BOBROW, S. A., & BOWER, G. H. Comprehension and recall of sentences. *Journal of Experimental Psychology*, 1969, **80**, 455–461.

BOWER, G. H. Mental imagery and associative learning. In L. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, in press.

HOLLAND, J. G., & KEMP, F. D. A measure of programing in teaching-machine material. *Journal of Educational Psychology*, 1965, **56**, 264–269.

KEMP, F. D., & HOLLAND, J. G. Blackout ratio and overt responses in programed instruction: Resolution of disparate results. *Journal of Educational Psychology*, 1966, **57**, 109–114.

PAIVIO, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–263.

STOLUROW, L. M., & WALKER, C. C. A comparison of overt and covert response in programmed learning. *Journal of Educational Research*, 1962, **55**, 421–429.

TOBIAS, S., & WEINER, M. Effect of response mode on immediate and delayed recall from programed materials. *Journal of Programed Instruction*, 1963, **2**, 9–13.

WICKELGREN, W. A. Acoustic similarity and retroactive interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1965, **70**, 102–108.

WICKELGREN, W. A. Phonetic similarity and interference in short-term memory for single letters. *Journal of Experimental Psychology*, 1966, **71**, 396–404.

(Received June 29, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 5, 400-404

EFFECTS OF VARIATIONS IN CONTENT AND STIMULUS GROUPING ON VISUAL SEQUENTIAL MEMORY[1]

JOHN R. BERGAN,[2] BARRY J. ZIMMERMAN, AND MAUREEN FERG
University of Arizona Center for Early Childhood Education

The influence of variations in content and stimulus grouping on sequential memory was studied in 103 fifth-grade, white Anglo children from lower and middle social classes. Three types of stimulus content were used: numerical, figural, and verbal. Single, double, and triple stimuli presentations were involved within each content dimension. Achievement and intelligence test data for each subject were included. Factor analyses revealed two stimulus grouping factors which were labeled single stimulus and multiple stimuli groupings. The results were discussed with regard to Jensen's research. It was concluded that what Jensen termed associative learning is determined not only by such factors as memory storage capacity but but also by the manner in which stimuli are grouped for recall.

In recent years, in large measure because of the research of Arthur Jensen, there has been increased recognition of the potential importance of sequential memory as an indicator of basic learning ability. Jensen has argued that the digit span test, which can be described as a measure of sequential memory, assesses basic, genetically determined, associative learning ability. Associative learning involves "neural registration and consolidation of stimulus inputs and the formation of associations [Jensen, 1969, p. 110]."

Associative learning is potentially important in education in a number of ways. First, according to Jensen (1969) associative learning appears to be a necessary, although not sufficient, condition for the development of higher order cognitive learning skills such as problem solving and concept learning. Second, associative learning skill is purported to be relatively independent of social class. Finally, insofar as associative learning is independent of social class, Jensen argues that the associative learning process could provide a basis for the establishment of instructional procedures which might enhance the educational attainments of disadvantaged children.

In light of the potential influence in education of associative learning, it is important to determine those variables which affect associative learning ability. This study investigates the effects of two variables, stimulus grouping and stimulus content, on sequential memory ability. Sequential memory is defined by tasks in which familiar stimuli are presented serially. The subject's task is to recall the order of presentation. In this study, stimuli were presented visually. All of the stimuli had known verbal labels which were used in recording presentation order.

Stimulus content refers to culturally determined stimulus categories. One of the most persistent findings in research on human abilities is that variations in content reduce relationships among abilities measuring the same operation or process. In Guilford's (1967) extensive investigations, for example, it has been observed repeatedly that the ability to perform a given intellectual operation does not generalize across content categories.

One possible explanation for the influence of content variations on ability is that people differ with respect to the amount and kinds of experience which they have had vis-à-vis

---

[1] This investigation was supported by the Follow Through component of the Arizona Center for Early Childhood Education, Ronald W. Henderson, Director, and was performed pursuant to a contract with the United States Office of Education, Department of Health, Education and Welfare. The authors would like to express their gratitude to Elaine Nicholson and Maure Hurt, Jr. for their assistance in the data analysis and to the teachers and administrators of Amphitheatre School District, Tucson, Arizona for their cooperation in the conduct of the study.

[2] Requests for reprints should be sent to John R. Bergan, Center for Early Childhood Education, College of Education, University of Arizona, 1515 East First Street, Tucson, Arizona 85719.

various content categories. Such differences may be common within ethnic groups and social classes. Lesser, Fifer, and Clark (1965) observed variation in verbal, numerical, and spatial ability patterns across ethnic groups. Such differences may be experientially based.

It was hypothesized that factor-analytic procedures would reveal separate factors reflecting variations in stimulus content. Three contents were used: words, numbers, and familiar shapes. Content variations were selected which, presumably, would not alter the basic character of the sequential memory task. All stimuli, regardless of content, were familiar to subjects and all stimuli had known verbal labels. If the variations in ability were to be produced by alterations in content, such variations would be a function of the type of stimulus being recalled rather than whether or not the stimulus was familiar or could be labeled.

Stimulus grouping refers to the clustering of stimuli on the basis of common concrete or abstract properties. For example, stimuli may be grouped in terms of such conceptual classifications as modes of transportation, tools, and so forth, or in terms of concrete properties such as contiguous occurrence in space or in time. Stimulus grouping in this investigation was defined as the number of stimuli presented simultaneously to the subject. Three stimulus grouping variations were used. Stimuli were presented one at a time, two at a time, or three at a time.

A number of experiments have indicated that variations in stimulus grouping may affect recall. Miller (1956) cites a series of investigations which demonstrated that when bits of information can be conceptually grouped into chunks by recoding input in terms of preestablished categories, recall can be increased. For example, recoding binary digits into the octal system greatly increased the number of binary digits which could be recalled in a sequential memory task. Jensen (1969) found that when words in a free-recall task represented a limited number of conceptual categories (e.g., food and clothing), recall was substantially better than when words did not lend themselves to conceptual grouping. However, improved performance was class related. Middle-class subjects improved greatly whereas the improvement of lower-class subjects was minimal. The ability to group into conceptual categories was also found to be highly related to IQ. Jensen concluded that recall involving conceptual groupings was a measure of what has traditionally been called general intelligence, an ability distinct from associative learning.

Before 1900, there was evidence of the influence on recall of grouping based on concrete properties. Müller and Schumann (1894) observed that subjects tended to group stimuli into rhythmic patterns and that attempts to suppress the use of rhythm decreased learning speed. These findings have been corroborated in a number of later studies (e.g., McGeoch & Irion, 1952). As indicated by the above research, the effects of stimulus grouping on mean level of performance in memory tasks are well established. However, there is no research indicating that stimulus grouping produces differential ordering of subjects in sequential memory ability.

The present study investigated the effects of grouping on individual differences in performance. It was hypothesized that variations in stimulus grouping would produce variations in the rank order of subjects in sequential memory performance. This hypothesis was based on the assumption that the group patterns used in stimulus presentation would influence individual variation in strategies used for recall. When stimuli were presented singly, the subject would be free to group the stimuli rhythmically in a variety of ways through covert rehearsal. When stimuli were presented in groups of two or three, it was assumed that these groupings would be used in recall.

METHOD

Subjects

Forty-three boys and 60 girls from fifth-grade classes drawn from three elementary schools in Tucson, Arizona, served as subjects. The children were predominantly white Anglo from either lower or middle socioeconomic groups.

Tests

The sequential memory task in all conditions of the experiment was to recall a series of stimuli presented visually at regular intervals by slide projector. The subjects responded after each series by signifying the stimuli in the series in writing on an answer sheet. White stimuli on a dark background were presented on a 6 × 4 foot screen by a 35-millimeter slide projector located approximately 30 feet from the subjects. The presentation of the slides was standardized by automating the triggering mechanism of a Kodak Carousel slide projector with a Lafayette electric timer. Each slide was presented for 2 seconds duration with a 1-second lapse interval between slides. After each trial, the experimenter manually halted the projector for 10 seconds to permit each child sufficient time to record his response on provided answer sheets. The answer sheet was composed of horizontally adjacent boxes, identical in number to the number of slides in each respective series. These rows of boxes were lettered successively. The experimenter verbally specified each row letter prior to the presentation of the correlative series.

Stimuli from three content dimensions were involved: numerical, figural, and verbal. Numerical content was comprised of numbers ranging from one to three; figural content involved circles, triangles, and squares; verbal content consisted of three words: "the," "loop," and "rope." Stimuli from the three content dimensions were recorded in the appropriate boxes in the following manner: For numerical stimuli, the subjects was instructed to insert the arabic numeral 1, 2, or 3. For figural stimuli, the subject was directed to draw a circle, triangle, or square. For verbal stimuli, the subject was instructed to insert the first letter of each word, i.e., (t)he, (l)oop, (r)ope.

Three modes of presentation were involved in each content dimension: the single stimulus (SS), the double stimuli (DS), and the triple stimuli (TS) presentation.

The SS presentation consisted of successively displaying a single stimulus on the screen. This was accomplished by placing a single stimulus (word, number, or shape) on each slide. Pilot testing revealed that fifth-grade children had no difficulty with series of four slides. Two sample series three and four slides in length were used to familiarize the youngsters with the task and to provide the experimenter with cues indicating whether the children understood the task. The first test series was composed of four successively presented slides. Succeeding series were gradually incremented by one slide; the final series reached nine slides in length. There were four different slide sequences at each series length. Considered together, 24 separate SS series were presented for each content dimension.

The DS presentation involved the successive display of two horizontally adjacent stimuli on the screen. This was effected by putting two stimuli on each slide. Pilot testing suggested that fifth-grade youngsters had no difficulty with two DS slide spans. Two sample series of one and two DS slides were used for familiarization and feedback purposes. The first test series was comprised of two successively presented DS slides. Succeeding series gradually incremented by one slide so that the final series reached seven slides in length. Thus, 24 separate DS spans were presented for each content dimension.

The TS presentation was composed of the successive display of three horizontally adjacent stimuli on the screen. This was accomplished by placing three horizontally adjacent stimuli on each slide. Pilot testing revealed that fifth graders had no difficulty with one TS slide series. Two sample series of one and two TS slides were used for familiarization and feedback purposes. The first test series was composed of a single TS slide. Series length was gradually increased by one slide until it totaled five TS slides. There were two different TS slide sequences at the single slide series length. From the two slide series length to the five slide series length, six different TS slide sequences were involved at each series length. In all, 26 separate TS series were presented in each content dimension.

Procedure

The memory tasks were presented to an entire classroom unit in the classroom itself or in the school multipurpose room. Random procedures were used to determine the presentation sequence for the nine subtasks. Only three subtasks were presented on a single day. The subtasks were generally administered on 3 successive days; however, scheduling difficulties sometimes necessitated that testing be delayed for a day or two.

The tasks were administered by a single experimenter who was assisted by two graduate student proctors. The experimenter verbally presented the instructions about the task and monitored the slide projector. The proctors were responsible for checking that each child was marking the appropriate series on his answer sheet and for supplying needed materials, for example, pencils. All subjects were directed to hold their pencils in the air during the presentation of each series and to mark their answer sheets only after the entire series was presented. This procedure enabled the proctors to monitor the subjects without difficulty. Proctors collected the answer sheets immediately upon completion of each subtask.

All protocols were hand-scored by graduate students. The scoring procedure paralleled that suggested by Jensen (1964). Each box in each series was scored as being either correct or incorrect. Correctly marked boxes were then summed across each series and then totaled for each subtask. In all, 1,890 separate boxes were involved in the nine subtasks.

Because of the potential educational importance of associative learning it was decided to include IQ and achievement data in the analysis. Achievement and intelligence test data were collected from school district records for each subject. The Stanford Achievement Test and the Otis Lennon Intelligence Test were administered within 3 months prior to the outset of the study.

TABLE 1

INTERCORRELATIONS OF SEQUENTIAL MEMORY SUBTESTS, INTELLIGENCE, AND ACHIEVEMENT

Test	A	V/SS	N/SS	F/SS	V/DS	N/DS	F/DS	V/TS	N/TS	F/TS
Intelligence	.83***	.31***	.35***	.32***	.33***	.37***	.28**	.37***	.45**	.23*
Achievement (A)		.22*	.38***	.30**	.20*	.38***	.22*	.35***	.40***	.22*
Verbal SS (V/SS)			.65***	.60***	.36***	.38***	.36***	.31***	.35***	.16
Numerical SS (N/SS)				.59***	.22*	.39***	.39***	.31***	.31***	.18
Figural SS (F/SS)					.55***	.63***	.46***	.43***	.50***	.35***
Verbal DS (V/DS)						.60***	.50***	.61***	.64***	.56***
Numerical DS (N/DS)							.56***	.52***	.59***	.48***
Figural DS (F/DS)								.60***	.49***	.51***
Verbal TS (V/TS)									.70***	.71***
Numerical TS (N/TS)										.68***
Figural TS (F/TS)										

* $p < .05$.
** $p < .01$.
*** $p < .001$.

RESULTS AND DISCUSSION

Examination of the standardized test scores revealed the group to be a little above average in both intelligence ($M = 108.6$, $SD = 12.9$) and achievement (grade level $M = 5.8$, $SD = 1.4$). The nine memory variables and the IQ scores were factor analyzed using principal axis and Varimax rotation procedures (Harmon, 1960).

Table 1 presents intercorrelations among memory variables and IQ. Correlations of memory subtest scores with IQ range from .18 to .45. The majority of the coefficients are in the .30s indicating that across variations in stimulus content and in stimulus grouping there is a significant relationship between memory for sequences and intelligence and achievement test performance.

Table 2 presents rotated factor loadings from the factor analysis. Of the variance among the test intercorrelations, 74% was accounted for by the three factors which emerged in the analysis. Factor I was labeled memory for multiple units of information, Factor II, memory for single units of information, and Factor III, intelligence and achievement. Results of the factor analysis give clear support to the hypothesis that memory for sequences involving groups of stimuli represents a separate ability from memory for sequences of single stimuli. These results augment the well-established findings that variations in stimulus grouping affect mean level of performance. Variations in stimulus grouping during presentation differentially influence the ordering of subjects in memory ability as well as their average level of performance. Furthermore the communalities for the various memory subtests indicate that the tests revealing these differential influences are reliable.

When stimuli are presented singly, there is nothing to prevent subjects from grouping them rhythmically through subvocal rehearsal. The possibility of grouping provides subjects with a potential strategy for enhancing their sequential memory ability. When stimuli are already grouped, as was the case for tasks involving simultaneous presentation of two or three stimuli, the possibility of using the grouping strategy is minimized.

Results of the factor analysis did not support the hypothesized emergence of factors representing differences in stimulus content. Factor loadings across content categories were quite high.

The central conclusion to be gleaned from

TABLE 2

ROTATED FACTOR LOADINGS OF SEQUENTIAL MEMORY SUBTESTS, INTELLIGENCE, AND ACHIEVEMENT FOR ALL SUBJECTS

Test	I	II	III	h^2
Intelligence	.21	.20	*.90*	.89
Achievement	.16	.16	*.93*	.92
Verbal SS	.15	*.85*	.10	.76
Numerical SS	.09	*.83*	.24	.76
Figural SS	.43	*.74*	.10	.74
Verbal DS	*.77*	.26	.05	.67
Numerical DS	*.64*	.43	.19	.62
Figural DS	*.66*	.36	.04	.57
Verbal TS	*.83*	.14	.22	.75
Numerical TS	*.79*	.19	.30	.75
Figural TS	*.87*	.02	.09	.76

Note.—Italicized numbers indicate loadings above .50.

the present study is that sequential memory ability as well as mean level of sequential memory performance is influenced by the manner in which stimuli are grouped. This suggests that what Jensen describes as associative learning ability is determined not only by such variables as capacity of the memory storage system, but also by the influence of the manner in which stimuli are grouped for recall.

Jensen (1969) has argued that conceptual grouping and associative learning are both genetically determined abilities. Certainly the conceptual grouping strategy is difficult to teach because it requires the learning of large numbers of concepts. On the other hand, one might guess that concrete grouping strategies would be easy to teach. A next logical step in research on sequential memory as a measure of basic genetically determined learning ability would be to investigate the influence of instruction in grouping through subvocal rehearsal, not on mean level of performance, but on the ordering of subjects in sequential memory ability.

REFERENCES

GUILFORD, J. P. *The nature of human intelligence.* New York: McGraw-Hill, 1967.

HARMON, H. H. *Modern factor analysis.* Chicago: University of Chicago Press, 1960.

JENSEN, A. R. *Individual differences in learning: Interference factor.* (United States Office of Education Cooperative Project No 1967.) Berkeley: University of California, 1964.

JENSEN, A. R. How much can we boost I.Q. and scholastic achievement? *Harvard Educational Review*, 1969, 39, 1–123.

LESSER, G. S., FIFER, G., & CLARK, D. H. Mental abilities of children from different social-class and cultural groups. *Monographs of Social Research in Child Development*, 1965, Serial No. 102, 30, No. 4.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.

MILLER, G. A. The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81–97.

MÜLLER, G. E., & Schuman, F. Experimentalle Beitrage zur Unterschuchung des gedachtnessess. *Zeitschrift für Psychologie.* 1894, 6, 81–190.

(Received July 8, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 5, 405–409

COMPARATIVE STUDY OF THE PERCEIVED RELEVANCE OF MATERIAL TO BE LEARNED AND ITS IMPACT ON THE PERFORMANCE OF CULTURALLY DEPRIVED JUNIOR COLLEGE STUDENTS

LAWRENCE R. O'LEARY[1]
University of Missouri, St. Louis

The relevance of material to be learned and its impact upon learning was investigated using the learning performance of 96 junior college students from different races, social classes, and ability levels. An idiographic method was utilized assessing the relevance of a given learning unit to a specific subject. As predicted, the improvement in performance that accompanies the introduction of more personally relevant material was greater for the Negroes than for the Caucasians. Similarly, the performance of low-ability subjects was more enhanced than that of high-ability students when personally relevant material was to be learned. Results are important for the education of the culturally deprived and low-ability students who constitute the most difficult education problem in contemporary society.

The important distinction between denotative (explicit) meanings and connotative (implicit) meanings is seriously distorted when psychologists attempt to investigate the effect of meaning through the use of consonant-vowel-consonant trigrams. The traditional tack in researching this area has been to use the concept of meaningfulness of consonant-vowel-consonant trigrams to measure indirectly the effect of meaning on material to be learned. A failure of previous studies concerned with meaningfulness has been the apparent reluctance to recognize, in a methodic manner, the uniquely individual aspect of meaning. A method of idiographically evaluating both the frequency dimension (association value) and intensity dimension (reinforcement value) has been devised by Rychlak (1966), who not only demonstrated the validity of the method but also the existence of an intensity continuum as an important dimension independent of association value.

The culturally deprived background of the Negro and the poor white results in an individual apperceptive mass that is greatly reduced or at least substantially different from the materially more fortunate members of our society. This intensity dimension of meaningfulness reflects the "relevance" of a given verbal unit to the individual. One hypothesis of the present study predicts an increase in the learning performance of both culturally advantaged and culturally disadvantaged when "relevance" is increased. More specifically, and more central to this study, it is predicted that the rate of increase in learning will be greater for the latter group than for the former.

The relevance approach contends that members of the culturally advantaged groups possess a more extensive vocabulary as well as an increased generalized desire to learn than subjects categorized as culturally disadvantaged. The frequency measure of meaningfulness, which assesses the number of meaningful ties in the standing language repertoire of an individual, is actually more relevant to a culturally enriched group. The disadvantaged person relies on his more uniquely held affective dimension of meaningfulness due to the inadequacies of his "programmed" verbal contents. The present study attempts to specify and investigate the dimension of relevance and the importance of its effect upon the learning of the culturally deprived.

---

[1] The author would like to acknowledge the substantial contribution of ideas and effort made by his wife Gerty in completing much of the work on which this article is based.

Requests for reprints should be sent to L. R. O'Leary, Evening College, University of Missouri, 8001 Natural Bridge Road, St. Louis, Missouri 63121.

METHOD

Eight hundred junior college students were asked to rate 149 consonant-vowel-consonant trigrams from a 65% to 75% range of association values in Archer's (1966) list. Their ratings were on two dimensions and were given on three separate occasions. On the first date, the students were asked to indicate which of the 149 trigrams reminded them of a word. Approximately 1 week later, the same students were asked to rate the identical 149 trigrams by indicating whether they had a positive or negative reaction to the unit (no matter how slight).

A 4-point rating scale was used to determine whether the students: (*a*) liked the unit a great deal, (*b*) liked it slightly, (*c*) disliked it slightly, or (*d*) disliked it a great deal. The likability rating was repeated 1 week later and the trigrams that were inconsistently rated in terms of likability, on the two separate occasions, were discarded. As a result of the ratings, each subject possessed a list of trigrams falling into four different categories. These included: (*a*) trigrams that he viewed as words and that he liked (W+), (*b*) trigrams that he viewed as words and that he did not like (W−), (*c*) trigrams that he did not think were words and that he liked (NW+), and, finally, (*d*) trigrams that he did not view as words and that he did not like (NW−). Therefore, each of these four types of trigram pairs (W+, W−, NW+ and NW−) had both a reinforcement value and an association value rating.

A mixed list of 16 consonant-vowel-consonant trigram pairs was generated for each subject in the study. Four of these pairs were from each of the above categories and both members of each pair came from the same category. The learning task was a modified paired-associate learning procedure. Any given pair was eliminated from the list after two consecutive associations and the number of trials to criteria was the score for a given pair.

Finally, means were computed to determine the number of trials required to learn liked material versus disliked material and trigrams having association value versus trigrams not having association value.

Performance on the learning task was then related to a number of organismic variables such as race, social class (cf. Hollingshead & Redlich, 1958), ability level (using total School and College Ability Test scores—long form), and sex.

RESULTS

Means and standard deviations were computed for each level of association value and reinforcement value. These reflect the mean number of trials required to learn that type of material. These statistics, computed for each of the organismic conditions considered in the present study, were computed and can be found in Table 1. The impact of the different dimensions of meaningfulness on the learning performance of the culturally deprived and low-ability subjects was examined through the use of an analysis of variance technique. However, since one organismic condition (high-ability Negroes) could not be filled, it was required that a series of analyses of variance be computed in an attempt to isolate the main effects and interactions bearing on the hypothesis. The original analysis examined the effect of five factors within the white sample (i.e., Socioeconomic Class × Ability × Sex × Reinforcement Value × Association Value; $n = 64$). Due to the absence of subjects in the Negro high-ability category, the race and ability factors cannot be examined in the same analysis. The results of the initial analysis of variance are found in Table 2.

The examination of the racial factor in addition to a desire to observe the impact of the likability (reinforcement value) and association value dimensions within the Negro group resulted in two additional analyses of variance. These were: (*a*) an analysis of variance between both races, all low ability, and (*b*) an analysis of variance within the Negro sample. Since each of the 96 subjects was measured a number of times on both the reinforcement value dimension and the association value dimension, the three previously mentioned analyses used an analysis of variance design with repeated measures for the last two factors as outlined in Winer (1962).

While there was no significant difference in the overall learning performance of members from the lower social classes and middle socioeconomic subjects, a significant F ($p <$ .01) reflected the superior learning performance of the Caucasians in this situation. Although high-ability subjects consistently surpassed low-ability subjects on the learning task, the differences in performance were not statistically significant.

The two remaining main effects of interest yielded some thought-provoking results. The main effect for association value ($p <$ .01) was obtained in both the white sample (high and low ability—see Table 2) and the combined racial group (all low ability—$p <$ .01). However, there was a failure of this

TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE NUMBER OF TRIALS NECESSARY FOR EACH GROUP TO LEARN THE FOUR TYPES OF MATERIAL

Group	Positive reinforcement value		Negative reinforcement value		Association value		No association value	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Socioeconomic class								
Middle	24.30	11.77	25.14	12.42	22.90	10.95	26.53	13.30
Low	23.06	7.71	23.89	8.35	22.24	8.02	24.70	8.04
Race								
White	24.16	10.33	25.64	9.88	22.24	10.32	27.56	11.30
Negro	31.34	13.37	37.45	15.64	33.53	14.38	35.26	14.64
Sex								
Male	27.38	13.08	29.51	14.58	27.42	13.52	29.47	14.13
Female	25.08	10.40	28.14	13.18	25.03	11.36	28.20	12.22
Ability								
High	23.18	9.53	23.39	11.04	23.40	10.27	23.67	10.30
Low	24.18	10.32	25.64	9.88	22.25	8.92	27.56	11.30

Note.—For the Sex subgroup $n = 48$; for all other subgroups, $n = 32$.

association value main effect to approach significance in the all-Negro sample. Interestingly enough, the impact of reinforcement value expressed itself in a significant *F* ($p = .01$) in both the all-Negro sample and the analysis combining the two races. However, a failure of the *F* to reach significance in the all-white sample (both high ability and low ability) apparently reflects the insignificance of the impact of changes in reinforcement value on the learning performance among high-ability subjects (see Table 2), a finding consistent with some previous reinforcement value research by Rychlak and Tobin (1971). This can be seen by examining the high-ability means for the two conditions of positive reinforcement and negative reinforcement in Table 1. There were no significant social class main effects.

It was predicted that Caucasian subjects would draw more on the frequency dimension of meaningfulness (association value) in their style of learning while the Negroes would draw more on the intensity dimension (reinforcement value) in their attainment of a learning criterion. While direct, statistically significant confirmation of this hypothesis was not forthcoming, there were some interesting trends in the predicted direction. These included a greater resultant increment in performance accompanying changes in association value in the white sample, while among the Negro subjects the greater increment accompanies changes in the level of reinforcement value.

Within the white sample, attention was given to the two indices of relevance (association value and reinforcement value) and their effect on the learning of low-ability and high-ability subjects. While reinforcement value had no differential effect on the learning of these two groups, the association-value aspect of meaningfulness had a significantly greater impact on the learning of low-ability subjects than on that of high-ability participants. This phenomenon was re-

TABLE 2

ANALYSIS OF VARIANCE OF TRIALS TO CRITERION FOR DIFFERENT LEVELS OF REINFORCEMENT VALUE, ASSOCIATION VALUE, AND A NUMBER OF ORGANISMIC VARIABLES, USING CAUCASIAN SUBJECTS

Source	*df*	*MS*	*F*
Between groups:	63		
Class	1	98.76	
Ability	1	166.56	
Sex	1	139.53	
Class × Ability	1	1,004.91	3.02
Error between	56	332.90	
Within groups:			
Reinforcement value (RV)	1	44.72	
Class × RV	1	0.00	
Ability × RV	1	26.45	
Sex × RV	1	57.42	1.14
Error within	56	50.24	
Association value (AV)	1	591.09	13.78*
Class × AV	1	41.97	
Ability × AV	1	331.61	7.73*
Sex × AV	1	2.08	
Error within	56	42.91	

* $p < .01$.

flected in an association value by ability interaction that was significant beyond the .01 level.

DISCUSSION

The major consideration motivating the present research was a belief that the use of an idiographically assessed relevance dimension would result in demonstrably greater increments in the learning performance of subjects from culturally deprived backgrounds, when compared to subjects in the middle class. Since membership in the Negro race is one substantial aspect of cultural deprivation in our present society, the significantly greater rate of improvement among the Negro population as compared with the Caucasian population possesses a number of implications. Among the Negro group, the mode of presentation or personal relevance of the material to be learned is a more crucial factor than among a white population. In other words, training programs as well as basic educational endeavors with culturally deprived Negroes should focus increased attention on the relevance of the material and of the presentation to the individual rather than maintain the traditional emphasis on the content to be presented.

Relevance has two basic components: a frequency (association value) and an intensity (reinforcement value) dimension. Obviously, the overall dimension of relevance and its impact on education has a cyclic quality. Increases in the familiarity (association value) of a given topic may result in increases in the intensity dimension toward that material to be learned. However, the thrust of the present results is the heightened importance of the intensity dimension in preparing educational and training programs directed primarily toward Negroes.

For example, a very definite answer has to be given to the question, "What's in it for me?" when presenting material to Negroes and very possibly to all minority groups. Their physical, intellectual, and emotive resources are directed toward developing a life style that is consistent with their past experience and frequently at odds with formal education. As a result, when asked to learn what 10 × 6 equals, for example, the basic ingredient in developing that bit of knowledge in the Negro is to show him why it is important to know, thus increasing the affective attitude that he has toward the material.

In terms of broad educational programs with Negroes the present results have certain implications.

1. Educational institutions dealing with underprivileged Negroes should have wide latitude in developing curricula that translate the content of primary and secondary school courses into digestible and relevant form.

2. This calls for research programs into the area of developing educational materials that not only have pictures of Negroes in the reading material, for example (although this is important), but also developing a method for measuring the personal relevance of different types of material to a given Negro student.

It should be noted, in reviewing the results, that the Negroes were still slower than Caucasians in learning material that they indicated has positive reinforcement value. This average superiority of approximately seven trials may be due to a number of factors that will be investigated in subsequent research. Among the more obvious explanations are: (*a*) the experimenter's race: Caucasian; (*b*) the use of written material (consonant-vowel-consonant trigrams) in the learning task; and (*c*) asking the subject to pronounce the trigram (when Caucasian phonics is one of the areas where Negroes differ from Caucasians most markedly).

Another interesting finding of this study was the relatively important aspect of association value on low-ability subjects. It is difficult to see specifically how association value translates into the everyday world of education. However, it does seem clear that a crucial aspect to be considered in educating children of low ability is the presentation of material in such a way that it can be directly related to previous experiences and information that the child possesses. Again, these findings call for increased flexibility of local schools to determine the nature of the curriculum for their student body, if not in the content then surely in the style of presenting the material.

REFERENCES

ARCHER, E. J. Re-evaluation of the meaningfulness of all possible trigrams. *Psychological Monographs*, 1960, 74(10, Whole No. 497).

HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness.* New York: Wiley, 1958.

RYCHLAK, J. F. Reinforcement value; a suggested idiographic intensity dimension of meaningfulness for the personality theorist. *Journal of Personality*, 1966, **34**, 311-335.

RYCHLAK, J. F., & TOBIN, T. J. Order effects in the affective learning styles of over achievers and underachievers. *Journal of Educational Psychology*, 1971, **62**, 141-147.

SKINNER, B. F. *The behavior of organisms: An experimental analysis.* New York: Appleton-Century, 1938.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received July 9, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 5, 410–412

EMPIRICAL TEST OF THE "TEACHER-CENTERED" VERSUS "STUDENT-CENTERED" DICHOTOMY[1]

FRANK COSTIN[2]
University of Illinois

An instrument was developed consisting of statements describing various kinds of classroom behavior to which college students could respond by judging how frequently the behavior had occurred. The statements were derived from characteristics which studies of college teaching and learning had classified as either "teacher-centered" or "student-centered." The instrument was administered to 626 college students in discussion sections of psychology and political science courses; these classes were taught by 23 graduate assistants. Principal axis analyses, followed by Varimax rotation, resulted in four distinct factors: (*a*) student involvement, (*b*) teacher support, (*c*) negative affect, and (*d*) teacher control. Intercorrelations among the factor scores supported the meaningfulness of the factors. Stability coefficients of factor scores for 219 additional students ranged from .67 to .77.

Some ten years ago, Richard Anderson (1959) carried out an extensive and critical review of studies concerning the relative effects of "authoritarian" and "democratic" leadership on learning.[3] He used these terms broadly, including studies which designated this kind of dichotomy with other labels, as, for example, "teacher-centered" versus "student-centered," "directive" versus "nondirective," "leader-centered" versus "group-centered." Anderson found that neither the authoritarian nor the democratic style of leadership was consistently related to greater production of learning. He also concluded that the authoritarian-democratic construct was a "grossly oversimplified" dimension, and inadequate for measuring leadership behavior.

Among the studies which led Anderson to his conclusion were those dealing with teacher-centered and student-centered discussion in the college classroom. Nevertheless, subsequent reviews and practical applications of such studies have not tended to question seriously the validity of these constructs. (For example, see McKeachie, 1963, pp. 1133–1140; 1967, pp. 38–40; Walker & McKeachie, 1967, pp. 38–40). The present investigation, therefore, sought to discover whether the kinds of college classroom behavior previously described in such terms as teacher-centered or student-centered would reveal, through factor analysis, more precise and meaningful dimensions.

METHOD AND RESULTS

Using the studies of college teaching reviewed by Anderson (1959) and by McKeachie (1963, 1967), a large number of characteristics were gathered which, according to the various investigators, represented either teacher-centered or student-centered activities and attitudes. These characteristics were rephrased into statements of behavior to which students could respond by judging the frequency with which it had occurred in the classroom. In addition to the present investigator, a panel of 10 graduate students enrolled in a seminar on the teaching of psychology decided whether each statement was clearly and "behaviorly" phrased, and whether it represented accurately the original characteristic it was based on.

Thirty statements were finally retained which all the judges had agreed upon.

---

[1] This study was supported in part by Office of Education Grant OEG-5-70-0031 (508). Thanks are due to the many teaching assistants and students who cooperated in the investigation; and to Edward Lewis, Department of Political Science, University of Illinois at Urbana-Champaign, and Reesa Vawter, Department of Psychology, Northern Illinois University, who made it possible for a considerable amount of the data to be collected.

[2] Requests for reprints should be sent to Frank Costin, Department of Psychology, University of Illinois, Champaign, Illinois 61820.

[3] Recently reprinted in Clarizio, Craig, and Mehrens (1970, pp. 606–624).

Fifteen of these statements represented teacher-centered characteristics, while the other half reflected student-centered characteristics (e.g. "Direction of discussion was controlled by the teacher," and "Direction of discussion was controlled by the students.")

The 30 statements were then arranged in a random order under the heading "Survey of Classroom Behavior," and administered at the end of the semester to 201 students in discussion sections of an introductory psychology course and an introductory political science course at the University of Illinois, Urbana-Champaign. These sections were taught by nine graduate assistants, none of whom had seen the Survey prior to its administration.

TABLE 1

FACTOR LOADINGS OF ITEMS ON THE SURVEY OF CLASSROOM BEHAVIOR

Factors	Analysis 1	Analysis 2
Factor I. Student Involvement		
5. Students volunteered knowledge, opinions, or personal experiences.	.60	.73
9. Students interacted with each other.	.73	.76
12. There was interaction between students and teacher.	.58	.68
13. Students talked more than the teacher.	.78	.72
20. Direction of discussion was controlled by the students.	.75	.67
Factor II. Teacher Support		
7. The teacher encouraged discussion of students' erroneous statements as a way of correcting them.	.58	.53
10. The teacher asked students to help determine content of discussion.	.57	.66
14. The teacher asked students to help determine objectives of discussion.	.58	.66
15. The teacher praised student behavior.	.42	.42
16. The teacher encouraged students to express their knowledge, opinions, or personal experiences.	.59	.51
17. The teacher asked students to help determine how their achievement would be evaluated.	.52	.66
19. The teacher asked open-ended questions.	.56	.43
Factor III. Negative Affect		
2. Students failed to laugh, joke, smile, or show other signs of humor.	.59	.69
3. The teacher corrected or rejected students' statements without further discussion.	.63	.70
6. The teacher made it clear that students would have little choice in how their achievement would be evaluated.	.49	.49
18. Students failed to ask teacher for information, opinions, or personal experiences.	.56	.49
Factor IV. Teacher Control		
1. The teacher defined the objectives of discussion.	.53	.45
4. Direction of discussion was controlled by the teacher.	.61	.71
8. The teacher defined the content of discussion.	.57	.58
11. The teacher asked specific, drill-type questions.	.42	.55

Note.—Factor loadings for Analysis 1 were based on the responses of 201 students. Factor loadings for Analysis 2 were based on the responses of 425 additional students. Responses to statements were scored on a 5-point scale, from "almost always" to "almost never." The number of a statement indicates its order on the Survey for Analysis 2.

TABLE 2

INTERCORRELATION OF FACTOR SCORES ON THE SURVEY OF CLASSROOM BEHAVIOR

Students	Correlation of factor scores		
	II	III	IV
Psychology students			
First administration			
I	.28	−.23	−.10
II		−.33	.26
III			−.09
Second administration			
I	.30	−.26	−.11
II		−.35	.21
III			−.001
Psychology students and political science students			
I	.32	−.27	.14
II		−.30	.29
III			.01

Note.—Roman numerals indicate the following factors: I = Student Involvement (five items); II = Teacher Support (seven items); III = Negative Affect (four items); IV = Teacher Control (four items). See Table 1 for list of items. For Psychology students, $N = 219$; for Psychology students and political science students, $N = 425$.

Students were asked to respond anonymously to each of the 30 items on the Survey according to these directions:

> The following statements refer to the behavior of your discussion section teacher, or to the behavior of students in the class. Print the letter A, B, C, D, or E at the left of each statement to show how frequently you think this behavior occurred during the semester. Answer according to this code:
>
> A. Almost always
> B. Often
> C. Occasionally
> D. Not usually
> E. Almost never

A principal axes factor analysis was carried out; examination of the results suggested a subsequent Varimax rotation to four factors (Harman, 1967). Using a loading of .400 as an arbitrary minimum criterion for including an item in a factor, 20 statements were retained. These were arranged in a random order to form a new Survey, the directions being identical to those previously used in the 30-item instrument. The Survey was then administered at the end of another semester to 425 additional students in discussion sections of the psychology courses and the political science course. A total of 14 graduate assistants taught these sections, none of whom had been involved in the earlier analysis.

Responses to the 20 statements on the Survey were subjected to the same analytic procedures as previously described, with four factors emerging. Table 1 shows the names which the investigator considered appropriate for the factors, the statements used in the Survey, and the factor loadings resulting from both analyses.

The percentage of variance obtained for each factor in the second analysis was as follows: Factor I—*Student Involvement*, 44.1; Factor II—*Teacher Support*, 22.8; Factor III—*Negative Affect*, 19.4; Factor IV—*Teacher Control*, 16.7.

The 20-item Survey was then administered twice to the same 219 students in discussion sections of an introductory psychology course at Northern Illinois University; a total of 11 graduate assistants taught these sections. This procedure was carried out toward the end of the semester, with a 2-week interval between testings. Correlations between the factor scores on the two administrations were as follows: Factor I, .74; Factor II, .77; Factor III, .72; Factor IV, .67.

Table 2 shows the intercorrelations among factor scores for the students at Northern Illinois University and for the students at the University of Illinois.

DISCUSSION AND CONCLUSIONS

Four factors emerged from analyses of college students' classroom observations (Table 1). The Survey statements underlying Factor I and Factor II were based on characteristics which previous research on college teaching had lumped together into a single democratic (or student-centered) style; nevertheless, correlations between the scores on these two factors were low, ranging from .28 to .32. The statements underlying Factor III and Factor IV were based on characteristics which previous research had grouped into a single authoritarian (or teacher-centered) style; yet, correlations between the scores on these two factors were practically zero. Correlations between scores on Factor II and scores on Factor IV were positive, but sufficiently low (ranging from .21 to .29) to strengthen the inference that the two factors represented different dimensions of behavior.

All in all, the results of this study were consistent with a conclusion which Anderson reached after his review of research: the "authoritarian-democratic" construct (or its semantic equivalents) is an oversimplified and inadequate basis for measuring teaching style and student behavior in the college classroom. The Survey instrument described here, and the dimensions of classroom behavior which it revealed, seem to be a more promising approach, especially for studies involving cognitive or affective outcomes of college courses.

REFERENCES

ANDERSON, R. C. Learning in discussion: A resume of the authoritarian-democratic studies. *Harvard Educational Review*, 1959, 29, 201–215.

CLARIZIO, H. F., CRAIG, R. C., & MEHRENS, W. A. (Eds.), *Contemporary issues in educational psychology*. Boston: Allyn & Bacon, 1970.

HARMAN, H. *Modern factor analysis*. Chicago: University of Chicago Press, 1967.

MCKEACHIE, W. J. Research on teaching at the college and university level. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally, 1963.

MCKEACHIE, W. J. New developments in teaching: New dimensions in higher education. (Report No. 16, Contract No. OEC 2-6-061722-1742, United States Office of Education, Department of Health, Education, and Welfare, with Duke University, 1967.)

WALKER, E. L., & MCKEACHIE, W. J. *Some thoughts about teaching the beginning course in psychology*. Belmont, Calif.: Brooks/Cole, 1967.

(Received July 27, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 5, 413-420

EFFECTS OF MEMORY SUPPORT ON STATE ANXIETY AND PERFORMANCE IN COMPUTER-ASSISTED LEARNING[1]

BARBARA L. LEHERISSEY,[2] HAROLD F. O'NEIL, JR., AND DUNCAN N. HANSEN
Florida State University

The hypothesis that memory support (MS) reduces state anxiety (A-State) and errors in a computer-assisted learning task was investigated. It was predicted that high A-State students given memory support would make fewer errors than high A-State students who were given no memory support (NMS). Low A-State students were expected to perform equally as well with or without memory support. Sixty male undergraduates were randomly assigned to MS or NMS conditions. All students received the same learning materials. The MS group was allowed to see their previous incorrect responses to each problem before attempting it again, whereas this information was not available to the NMS group. There were no significant effects of memory support on A-State as measured by the State-Trait Anxiety Inventory. However, the predicted Memory Support × A-State interaction was found for errors. These findings were compared with the results of previous research on anxiety and memory support.

The purpose of the present study was to determine the relationship between state anxiety and memory support on a complex computer-assisted learning task for persons differing in trait anxiety. Hypotheses about the relationship between anxiety and performance in a learning task may be derived from the Drive Theory of Spence (1958) and Taylor (1956), which predicts that the effects on performance of individual differences in anxiety (drive) level will depend upon the relative strength of the correct response and competing error tendencies. High drive would be expected to facilitate performance on simple learning tasks where the correct response is dominant, and to produce debilitating effects on difficult tasks where error tendencies are stronger.

In research on Drive Theory, it is generally assumed that scores on the Taylor (1953) Manifest Anxiety Scale (TMAS) reflect individual differences in drive level. Spielberger (Spielberger, 1966; Spielberger, Lushene, & McAdoo, in press) has pointed out that the TMAS seems to measure trait anxiety, while the concept of drive is logically more closely associated with state anxiety. State anxiety (A-State) refers to a transitory state or condition that is characterized by feelings of tension and apprehension and heightened autonomic nervous system activity, whereas trait anxiety implies individual differences in anxiety proneness, that is, the disposition to respond to elevations in A-State under conditions that are characterized by some threat to self-esteem. It would be expected, therefore, that persons who differed in trait anxiety would manifest differences in drive level only under circumstances that caused them to respond with differential elevations in state anxiety.

A general paradigm for studying the effects of anxiety on cognitive activity has been presented by Sieber and Kameya (1967). This paradigm includes the following features:

> a) assessing initial level of anxiety, b) determining one or more mediating process variables (e.g., discrimination, short term memory, ideational fluency) and an overall performance measure (e.g., learning speed) that are affected by anxiety, and then c) measuring the degree to which anxious and non-anxious persons manifest change in those mediating process

---

[1] The authors wish to thank Charles D. Spielberger for assistance on an earlier draft of this manuscript, and Barbara Johnson for editorial comments. This research was supported by The Office of Naval Research, Contract No. N00014-68-A-0494, Project NR 154-280. A paper based on this research was presented at the 1970 American Educational Research Association Meeting, Minneapolis, Minnesota.

[2] Requests for reprints should be sent to Barbara L. Leherissey, Computer-Assisted Instruction Center, 1-A Tully Building, Florida State University, Tallahassee, Florida 32306.

> and overall performance variables following some experimental treatment [p. 2].

Sieber has examined the hypothesis that anxiety interferes with memory, but "when memory support is provided, anxious persons appear to take advantage of it and thereby to improve their level of performance [Sieber, 1969, p. 59]." Using the above paradigm, Sieber and Kameya (1967) investigated the relationship between test anxiety, as measured by the Test Anxiety Scale for Children (Sarason, Davidson, Waite, & Ruebush, 1960), and the effects of memory support on the problem-solving ability of children. On a Marble Task, high-anxious students performed as well as low-anxious students when given memory support, but without memory support, high-anxious students made more errors than low-anxious students. Low-anxious students performed equally as well with or without memory support. Sieber and Kameya (1967) conclude that anxiety interferes with short-term memory, but that the performance of high-anxious students can be improved when the task is structured to provide external aids so that there is less dependency upon memory.

Paulson (1969) used essentially the same paradigm as Sieber and Kameya, and found no differences in the error rate for high- and low-anxious students when memory support was provided, but that high-anxious students did more poorly without memory support. Since his concept identification task differed markedly from Sieber and Kameya's Marble Task, Paulson concluded that the facilitating effect of memory support for high-anxious students was not limited to a relatively narrow set of experimental tasks.

Several points can be raised regarding the theoretical interpretation and experimental design employed by Sieber and Kameya (1967) and Paulson (1969). First, Sieber (1969) argues that test anxiety is a process variable, and yet the paradigm she has chosen for investigating the effects of anxiety on memory requires only an initial assessment of anxiety level. If anxiety is regarded as a process variable, this would seem to imply a changing state of the organism in which anxiety level varies as a function of the individual's perception of the situation at a given point in time. Consistent with this view, measures of state anxiety should be taken at intervals within the experimental situation.

Sieber also contends that changes in performance rather than changes in level of anxiety are sufficient for investigating the effects of anxiety on intellectual activity. One of her major arguments for measuring performance changes and making only an initial assessment of anxiety level is based on the assumption that anxiety inventories are too insensitive to measure changes in anxiety level (Sieber, 1969). This may be true for anxiety measures such as the TASC, but Spielberger, et al., (in press) have argued that the A-State scale of the State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970) has demonstrated both construct and concurrent validity, and state that:

> While measures of A-Trait provide useful information regarding the probability that high levels of A-State will be aroused, the impact of any given situation on the intensity of A-State can only be ascertained by taking actual measurements of A-State in that situation.

The value of measuring State Anxiety in the experimental situation was demonstrated by O'Neil, Spielberger, and Hansen (1969) who investigated the effects of A-State on learning mathematical materials that were presented via computer-assisted instruction (CAI). In this study, the State-Trait Anxiety Inventory (STAI) (Spielberger, Gorsuch, & Lushene, 1970) was used to measure A-State during the learning task. High A-State students made more errors on the difficult portion of the learning task than low A-State students, but fewer errors on the easier portion of the task. In a follow-up study, O'Neil, Hansen, and Spielberger (1969) found essentially the same anxiety-task difficulty interaction. In neither study was level of A-Trait related to performance. Moreover, in the second study, the mean error rate for the high A-State students on the difficult portion of the task was approximately twice that of the first study.

The major difference between the two studies was that the learning materials were presented on a typewriter terminal (an IBM 1440 system) in the first study and on a cathode ray tube (CRT) (an IBM 1500 system) in the second. The typewriter terminals provided a printed output of learning materials and the students' responses, and thus, the student could review his previous erroneous responses prior to each trial. In contrast, the learning materials presented on a CRT were programmed to be erased immediately after the subject responded. Thus, there may have been a greater memory load with the CRT and this may have accounted for the higher mean error rates for high A-State students in the second CAI study. Apparently, the typewriter printout associated with the 1440 system provided greater memory support in the first CAI study, which facilitated performance and reduced error rates.

One method for reducing errors on learning materials presented on a CRT would be to provide some type of memory support. In the present study, a major goal was to investigate the effect of providing memory support on performance for persons differing in anxiety when presented complex learning materials on the CRT of an IBM 1500 system. It was expected that the findings of Sieber and Kameya (1967) and Paulson (1969) would be replicated. However, it should be recalled that an A-Trait measure was used in these studies, whereas A-State is logically more closely associated with drive level. Therefore, in the present study, the following hypotheses were formulated: (*a*) the performance of high A-State students would be inferior to that of low A-State students with out memory support; and (*b*) there would be no differences between high and low A-State students with memory support.

A second major goal of this study was to examine the effects of memory support on the A-State reactions of students who differed in A-Trait. Assuming that memory support reduces the amount of stress inherent in a learning task, it was expected that there would be less difference in the level A-State of high and low A-Trait students in the memory support condition than in the condition without memory support.

METHOD

Subjects. The students were 60 male undergraduate student volunteers at Florida State University who were enrolled in the introductory psychology class. The students were run in small groups of 8–15 students; a total of five experimental sessions was required to run all groups of students. On the basis of the order of their arrival for each experimental session, the students were randomly assigned to one of two experimental conditions, memory support (MS), or no memory support (NMS). Thus, in each experimental session, approximately half of the students were run in the MS condition and the other half were run in the NMS condition.

Apparatus. An IBM 1500 CAI system (IBM, 1967) was used to present the learning materials. The terminals for the system, which consist of a CRT, a light pen, and a typewriter keyboard, were located in a sound-deadened, air-conditioned room. The CAI system also presented the STAI A-State scales while students worked through the learning materials and recorded students' responses.

Learning materials and program description. The learning materials consisted of a mathematics program on the field properties of complex numbers which is reported in detail by O'Neil (1969). This program is divided into three sections (Parts A, B, and C) with five problems each, for a total of 15 problems. Each problem consists of a mathematics statement for which the student is required to select the most appropriate validating proof from a list of alternatives. The learning materials were programmed such that students were required to successfully solve each problem before being allowed to progress to the next problem.

Anxiety measures. The A-State and A-Trait scales of the STAI (Spielberger, Gorsuch, & Luschene, 1970) were used to measure anxiety. A total of five A-State measures were obtained during the experimental session: (*a*) The 20-item STAI A-State scale was administered on the CRT prior to presenting the CAI mathematics learning materials, and after program completion. (*b*) Three five-item short form A-State scales were inserted after each of the three sections of the learning materials. The short form A-State scales consisted of the five items having the highest item-remainder correlations for the normative sample of the STAI A-State scale (Spielberger, et al., 1970). Following the experimental session, the 20-item A-Trait scale was administered via paper and pencil.

The 20-item A-State scales were given with standard instructions, i.e., students were asked to indicate how they felt right now, at this moment. The three short form A-State scales were given with instructions which asked the students to indicate how they felt while they were working on the previous section of the mathematics learning task. The three short form A-State scales were programmed to provide random presentation of the five items during each of the three administrations of these scales.

Procedure. The experimental session consisted of the following three stages: (*a*) the Pretask stage, during which students received instructions on the operation of the CAI terminals and took the first 20-item A-State scale on the CAI system; (*b*) the Performance stage, in which the students learned the mathematics materials and responded to the short form A-State measures after each part of the task was completed; (*c*) the Posttask stage, during which students took the second 20-item A-State scale on the CAI system, and were then given the A-Trait scale via paper and pencil, and finally interviewed and debriefed. Each of these stages is further described below:

1. *Pretask stage.* Students were randomly assigned to MS and NMS conditions according to order of arrival at the CAI Center. They were then assigned to CAI terminals and asked to read instructions on the operation of CAI features such as the light pen and typewriter keyboard. The students then took the 20-item A-State scale on the CAI system.

2. *Performance stage.* All students worked through the same learning materials at their own rate. During this stage, the students received either MS or NMS based on their assignment to memory condition. Students in the MS condition were initially provided the following special instructions and procedures by the CAI system:

> As you work through the math problems, the answers you type in will be stored. If you make an error, you will be permitted to see your previous wrong responses so that you can avoid entering the same wrong response more than once. In the case of wrong responses, a statement will appear at the bottom of the screen which states, "Your previous answers were." Your previous incorrect responses will then appear and remain on the screen until you make a correct response.

In the NMS condition, the CRT was blank for 20 seconds, which was comparable to the time required for students in the MS condition to read the above introductory instructions. In the learning task, the students in the MS group were shown all their previous incorrect responses to each problem before attempting the problem again. The students in the NMS condition were also required to attempt each problem again; however, their previous incorrect responses were not available. Following completion of each section of the learning materials, the students were asked to respond to the short A-State scale.

3. *Posttask stage.* Upon completion of the third short A-State scale, students were asked to respond to the second 20-item STAI A-State scale on the CAI system. The 20-item A-Trait STAI measure was then administered by paper and pencil, and this was followed by a 3-minute structured interview on the nature of the experimental task. In addition, students were asked not to discuss the experiment with any of their classmates.

RESULTS

Effect of Experimental Conditions and Trait Anxiety on Errors

In order to determine the effects of memory conditions and trait anxiety on performance, a three-factor analysis of variance with repeated measures on the last factor was calculated. Independent variables were levels of A-Trait, high (HA), medium (MA), and low (LA), Memory Conditions (MS, NMS) and Periods (A, B, C). Mean errors per correct response was the dependent variable in this analysis. The cut-off scores for the high A-Trait (HA) and low A-Trait (LA) groups corresponded to the upper and lower quartiles of the published A-Trait worms for college undergraduate males (Spielberger et al., 1970). Medium A-Trait (MA) students were those students whose scores were in the middle portion of the normative A-Trait distribution. The number of students within A-Trait level and memory condition was unequal but proportional for each group. The mean A-Trait scores for the students in the MS condition were 36.0; mean A-Trait scores for the NMS group were 35.2 ($t = .412$, $df = 58$, *ns*). Thus, the MS and NMS groups were well matched for A-Trait.

Results of the analysis of variance on mean errors made by HA, MA, and LA students in the MS and NMS conditions in Periods A, B, and C of the learning task revealed a significant main effect of Periods ($F = 11.91$, $df = 1/108$, $p < .001$). Examination of the mean errors across the three periods indicated that all groups made more errors on Period A ($\bar{X} = 4.57$) than on either Period B ($\bar{X} = 2.49$) or Period C ($\bar{X} = 2.79$). No evidence was found that errors in the three periods of the learning task were influenced by level of A-Trait or experimental conditions.

The failure to find any relationship between A-Trait and errors in the present study was consistent with the results of O'Neil, Spielberger, and Hansen (1969), O'Neil, Hansen, and Spielberger (1969), and O'Neil (1969) in which A-Trait and errors were also unrelated. However, in the three previous studies, there was a significant relationship between A-State and errors; consequently, in the present study, the relationship between A-State and errors was evaluated for the three sections of the difficult task.

Effect of Experimental Conditions and State Anxiety on Errors

A similar three-factor analysis of variance was calculated to determine the effects of in-task state anxiety and memory conditions on errors. The independent variables in this analysis were levels of A-State (high, medium, low), memory conditions (MS, NMS), and periods (A, B, C). The students were divided into high, medium, and low A-State groups based on their summed A-State scores across Periods A, B, and C. This distribution of in-task A-State scores was ranked and divided into thirds ($n = 20$ for each group). The MS and NMS students were then separated out of this distribution yielding an unequal but proportional n in each group. The range of high A-State was 36–60; medium A-State scores ranged 27–36; the range of low A-State was 15–27. Mean errors per correct response was again the dependent variable in this analysis.

The means and standard deviations for errors made by high, medium, and low A-State students in the MS and NMS conditions on Periods A, B, and C are shown in Table 1.

The most important finding in the analysis of variance on these data was a significant interaction between A-State and Memory Conditions ($F = 3.51$, $df = 2/54$, $p < .05$). A plot of this interaction is shown in Figure 1, which indicates that high, medium, and low A-State students were differentially influenced by memory conditions. High A-State students in the NMS condition made approximately 1.75 times as many mean errors per correct response as high A-State students in the MS condition and a t test indicated that these differences were significant ($t = 2.38$, $df = 58$, $p < .025$). In addition, high A-State students in the MS condition made significantly more errors than low A-State students in the MS condition ($t = 2.03$, $df = 58$, $p < .025$). Medium and low A-State students were found to perform approximately as well with or without memory support.

The analysis of variance on mean errors made by high, medium, and low A-State students in MS and NMS conditions across Periods A, B, and C also indicated that high A-State students made more errors ($\bar{X} = 5.04$) than either medium A-State students ($\bar{X} = 2.54$) or low A-State students ($\bar{X} = 2.27$). This main effect of A-State was significant at the $p < .01$ level ($F = 7.22$, $df = 2/54$). In addition, the main effect of Periods was significant at the $p < .001$ level ($F = 11.86$, $df = 2/108$).

TABLE 1

MEAN ERRORS PER CORRECT RESPONSE FOR HIGH, MEDIUM, AND LOW A-STATE STUDENTS IN MEMORY SUPPORT (MS) AND NO MEMORY SUPPORT (NMS) CONDITIONS ON PERIODS A, B, AND C

Groups	Period A	Period B	Period C
All groups			
M	4.57	2.49	2.79
SD	3.49	2.94	4.17
MS			
High			
M	4.26	3.67	4.24
SD	2.85	3.98	5.56
Medium			
M	3.30	1.92	1.52
SD	2.58	2.33	2.07
Low			
M	3.90	1.87	1.43
SD	2.47	1.81	1.34
NMS			
High			
M	9.63	4.23	8.13
SD	4.72	2.50	6.62
Medium			
M	4.14	2.10	2.24
SD	2.51	3.20	2.64
Low			
M	4.21	1.51	.91
SD	3.70	2.09	1.10

Note.—$N = 60$ for all groups.

Effects of Experimental Conditions on A-State for HA, MA, and LA Students

To investigate the relationships between levels of A-Trait and experimental conditions on A-State scores across the five periods, a three-factor analysis of variance with repeated measures on the last factor was calculated. Independent variables in this analysis were levels of A-Trait (HA, MA, LA), Memory Conditions (MS, NMS), and Periods (Pre, A, B, C, Post). The dependent measure was mean A-State scores in each of the five periods. The statistical analysis of the A-State data was based on the short form of the STAI A-State scale. The same five items in the short form were extracted from the total A-State scale given before and after the experimental task. The alpha reliabilities for the five A-State

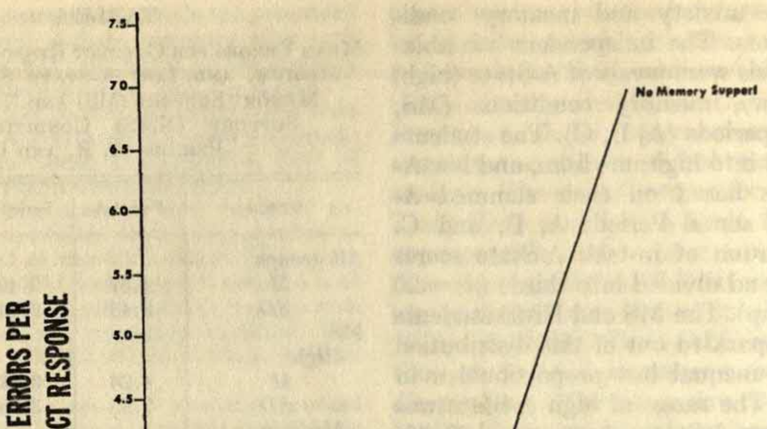


FIG. 1. Mean errors per correct response for low, medium, and high A-State subjects in the Memory Support and No Memory Support conditions.

scales were, respectively, .87, .83, .87, .86, and .93.

The means and standard deviations of A-State scores for HA, MA, and LA, students in the MS and NMS conditions for the five periods are reported in Table 2.

Results of the analysis of variance on these data indicated that HA students had higher A-State scores ($\bar{X} = 12.27$) than either MA ($\bar{X} = 9.75$) or LA ($\bar{X} = 8.50$) students. This main effect of A-Trait was significant at the $p < .001$ level ($F = 10.52$, $df = 2/54$). Students were also found to differ in mean A-State scores across the five periods ($F = 21.23$, $df = 4/216$, $p < .001$). The mean A-State scores for the MS condition ($\bar{X} = 10.39$) were higher than the NMS condition ($\bar{X} = 9.33$); however, this difference was not significant at the $p < .05$ level ($F = 2.90$, $df = 1/54$, $p < .10$). No other effects were observed.

DISCUSSION

The Sieber and Kameya (1967) and Paulson (1969) finding that high A-Trait students in the MS condition made fewer errors than high A-Trait students in the NMS condition was not supported in the present study. In this study, high, medium, and low A-Trait students were not found to differ significantly in mean errors for either MS or NMS conditions. The A-Trait results were consistent with the two previous CAI studies reported (O'Neil, Hansen, & Spielberger, 1969; O'Neil, Spielberger, & Hansen, 1969).

The predicted relationship was found, however, between A-State, Memory Conditions, and errors. Memory support was found to reduce the errors made by high A-State students, that is, high A-State students in the MS condition made fewer errors than high A-State students in the NMS condition, whereas medium and low A-State students performed equally as well with or without memory support. This significant interaction between levels of A-State and memory conditions may be interpreted as an indication that the performance of high state-anxious students can be facilitated by the provision of some type of memory support. This interpretation is consistent with that of previous memory support studies (Sieber & Kameya, 1967; Paulson, 1969) which point out the need for considering the differential effects of memory support for high-anxious versus low-anxious persons.

One interpretation for the superior performance of high A-State students in the MS versus the NMS condition may be derived from Drive Theory. Drive Theory (Spence & Spence, 1966) predicts that the effects of drive (state anxiety) on learning depend on the relative strength of correct and competing error tendencies. If, as Sieber and Kameya (1967) and Paulson (1969) claim, memory support reduces the disruptive effects of anxiety on memory, memory support may be seen as reducing the competing response tendencies, i.e., task-irrelevant responses, for high A-State students which would reduce their errors. Thus, superior performance of high A-State students in the MS as compared with the NMS condition would be expected.

The fact that Sieber and Kameya (1967) and Paulson (1969) used a trait rather than a state measure of anxiety and found a relationship with memory conditions and errors may have been due to the use of the TASC scale. This scale may be tapping anxiety which is more closely related to the situation than the STAI A-Trait scale used in the present study. As Spielberger et al. (in press) have stressed, inventories which measure anxiety aroused in a specific situation are more likely to be related to the behavior of students in that situation.

TABLE 2

MEAN A-STATE SCORES FOR HA, MA, AND LA STUDENTS IN MEMORY SUPPORT (MS) AND NO MEMORY SUPPORT (NMS) CONDITIONS ON PERIODS PRE, A, B, C, AND POST

Groups	Pre	Period A	Period B	Period C	Post
All groups					
M	8.25	11.65	10.22	10.05	9.13
SD	2.57	3.62	3.52	3.80	3.25
MS					
HA					
M	10.75	14.25	13.25	13.25	11.87
SD	1.83	3.69	3.01	2.82	4.22
MA					
M	8.27	12.36	11.55	10.73	10.55
SD	1.90	2.73	1.57	3.00	2.66
LA					
M	6.91	10.36	9.36	8.91	6.64
SD	1.97	4.08	4.32	4.48	2.38
NMS					
HA					
M	10.50	13.50	11.50	12.17	11.00
SD	3.15	3.51	2.26	3.43	2.53
MA					
M	8.17	10.67	8.83	8.58	8.17
SD	2.52	2.64	2.55	3.09	2.92
LA					
M	6.75	10.50	8.50	8.75	8.33
SD	2.01	3.94	4.17	3.91	2.23

Note.—$N = 60$ for all groups.

An interesting finding in the present study is the mean difference in A-State scores in the MS and NMS conditions. A-State scores were found to be higher for students in the MS group. Although this difference only approached significance ($p < .10$), it may be possible to infer that the MS condition was operating as a stress condition, in that providing these students with their previous incorrect responses was also providing them with a constant reminder of their past failures.

Also of interest in the present study was whether the provision of memory support would reduce errors of high A-State students on the 1500 system to the level of errors made by high A-State students on the 1440 system. The results of this study indicated that high A-State students in the MS condition made approximately as many errors as subjects on the 1440 system. Thus, it would appear for high A-State males that memory support was successful in reducing their errors on the CAI 1500 system.

In summary, the findings of the present study seem to indicate that the provision of memory support can improve the performance of male college students with high levels of state anxiety. However, the results of this study suggest that it may not be memory support per se that reduces the undesirable effects of anxiety on performance. A question of primary importance seems to be centered on the type of memory aid which is provided. The data reported seems to indicate that whereas providing high A-State students with their previous incorrect responses does facilitate performance, it also leads to an increase in levels of state anxiety for students in the MS condition. As has been suggested, threat of failure may be the undesirable side effect of memory support of this nature. A fruitful research effort would seem to be one which examines those memory aids which not only improve performance, but which reduce anxiety experienced in the learning situation.

REFERENCES

INTERNATIONAL BUSINESS MACHINES, INC. *IBM Coursewriter II author's guide. Part I: Course planning.* New York: IBM, 1967.

O'NEIL, H. F. Effect of stress on state anxiety and performance in computer-assisted learning. Unpublished doctoral dissertation, Florida State University, 1969.

O'NEIL, H. F., HANSEN, D. N., & SPIELBERGER, C. D. Errors and latency of response as a function of anxiety and task difficulty. Paper presented at the annual meeting of the American Education Research Association, Los Angeles, February, 1969.

O'NEIL, H. F., SPIELBERGER, C. D., & HANSEN, D. N. Effects of state-anxiety and task difficulty on computer-assisted learning. *Journal of Educational Psychology*, 1969, 60, 343–350.

PAULSON, F. L. The effect of memory support on the concept learning of test anxious children. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, February 1969.

SARASON, S. B., DAVIDSON, K. S., WAITE, R. R., & RUEBUSH, B. K. *Anxiety in elementary school children.* New York: Wiley, 1960.

SIEBER, J. E. A paradigm for experimental modification of the effects of test anxiety on cognitive processes. *American Educational Research Journal*, 1969, 6, 46–59.

SIEBER, J. E., & KAMEYA, L. I. The relation between anxiety and children's need for memory support in problem solving. (Revised Research Memorandum No. 11.) Stanford: Stanford Center for Research and Development in Testing, 1967.

SIEBER, J. E., KAMEYA, L. I., & PAULSON, L. Effect of memory support on the problem-solving ability of test-anxious children. *Journal of Educational Psychology*, 1970, 61, 159–168.

SPENCE, K. W. A theory of emotionally based drive (D) and its relationship to performance in simple learning situations. *American Psychologist*, 1958, 13, 131–141.

SPENCE, J. T., & SPENCE, K. W. The motivational components of manifest anxiety: Drive and drive stimuli. In C. D. Spielberger (Ed.), *Anxiety and behavior.* New York: Academic Press, 1966.

SPIELBERGER, C. D. Theory and research on anxiety. In C. D. Spielberger (Ed.), *Anxiety and behavior.* New York: Academic Press, 1966.

SPIELBERGER, C. D., GORSUCH, R. L., & LUSHENE, R. E. *Manual for the State-Trait Anxiety Inventory.* Palo Alto, Calif.: Consulting Psychologist Press, 1970.

SPIELBERGER, C. D., LUSHENE, R. E., & MCADOO, W. G. Theory and measurement of anxiety states. In R. B. Cattell (Ed.), *Handbook of modern personality theory.* Chicago: Aldine, in press.

TAYLOR, J. A. A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology*, 1953, 48, 285–290.

TAYLOR, J. A. Drive theory and manifest anxiety. *Psychological Bulletin*, 1956, 53, 303–320.

(Received July 27, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 5, 421-426

EFFECTS OF TRAINING BLACK PRESCHOOL CHILDREN IN VOCABULARY VERSUS SENTENCE CONSTRUCTION[1]

PAUL R. AMMON[2] AND MARY SUE AMMON
University of California, Berkeley

Seventy-two children were divided equally into vocabulary training, sentence training, and control groups. Each child was tested before and after a 6-week training period with the Peabody Picture Vocabulary Test (PPVT), a Sentence Imitation Test (SIT), and a Picture Interview (PI). Vocabulary training involved practice in recognizing and applying words from the PPVT and the PI. Sentence training consisted of imitating sentences similar to those in the SIT and constructing new instances of the same sentence types. Analyses of variance showed that vocabulary training had a positive effect on recognition and production of the words taught ($p < .025$), but did not transfer to sentence imitation. There were no significant effects of sentence training. Vocabulary seems more amenable to early training with disadvantaged children.

In recent years there has been considerable interest in facilitating the language development of young disadvantaged children through some sort of direct intervention. In spite of much activity in this area, very little is known about the differential effects produced by specific types of early language training. It is important to find out which aspects of language are most amenable to early training, and which types of training have the greatest transfer value for stimulating language development in general. Otherwise, attempts at intervention will be inefficient, or even inappropriate, with potentially damaging consequences for the children involved.

The present report is concerned with an experimental study which compared the effects of training black preschool children in vocabulary versus sentence construction. The rationale for this experiment was based, first of all, on the observation that previous studies have dealt with relatively complex language training programs, tending to confound the effects of different methods with those of different content (e.g., Cazden, 1965; Dickie, 1968). The strategy of the present study was to hold the method as constant as possible, while setting up a clear-cut difference in content. The specific contrast between vocabulary and syntax was suggested not only by a formal analysis of language, but also by some recent theorizing about language acquisition and the disadvantaged child. On the one hand, vocabulary has been singled out as the major area of language in which young children normally receive much deliberate instruction from adults (Cazden, 1968). Thus, if a child grows up around adults who do not pay much direct attention to his language, vocabulary may be the key to any language deficit which results. On the other hand, one popular belief nowadays is that disadvantaged children are brought up in a community which speaks mainly a "restricted code" (Bernstein, 1964). The restrictedness of the code is largely a matter of syntax, so that the child does not develop an ability to use the full range of syntactic possibilities for "elaborating" his sentences with words, phrases, and clauses which make meaning more explicit.

---

[1] This research was supported by the United States Office of Economic Opportunity (OEO Contract B99-4776). The present report was prepared at the University of California's Institute of Human Learning, which is supported by grants from the National Institutes of Health. An earlier version of this paper was presented at the Annual Meeting of the American Educational Research Association, Minneapolis, March 1970. The authors are grateful to Joel Levin, Nancy Suzuki, and Sylvia Zuck for their patient help in coding and analyzing the data. Sharon Alexander, Carol Meredith, Sharon Rose, and Caryl Sutton assisted in the testing and training phases of the research. Phyllis Koppelman and Pat Stohr transcribed much of the speech data.

[2] Requests for reprints should be sent to Paul R. Ammon, Department of Education, University of California, Berkeley, California 94720.

The available data on language abilities among disadvantaged black children are not very helpful in pinpointing the source of the deficiency which these children are reputed to have. They usually attain low scores on standard tests of vocabulary (Cazden, 1966; Raph, 1965), but these scores may simply reflect a symptom rather than a primary cause of deficiency. And although Bernstein's notion of a restricted code has often been used to characterize the language of lower-class black children (Bereiter & Engelmann, 1966; Deutsch, 1963; Osborn, 1968), it is not clear how legitimate this characterization is. There has been a tendency to interpret deviations from standard English as evidence of a restricted code. A distinction must be made between the features of a restricted code (as defined by Bernstein) and the features of a nonstandard dialect spoken in the black community. Labov (1970) and other sociolinguists have argued convincingly that Black English, defined as a system of grammatical rules, is no more restricted or impoverished than Standard English. Furthermore, it can be shown that black speakers produce some very complex speech, beginning at an early age. Linguistic analyses of this sort have led some writers to propose the use of a "difference model" rather than the more usual "deficit model" in comparing lower-class blacks to middle-class whites (e.g., Baratz, 1969). But while a difference model does seem more appropriate for many social class and ethnic comparisons, the deficit model may still have some validity. Disadvantaged black children may, in fact, be somewhat less skilled than their middle-class peers in using the syntactic and lexical resources of their own communities to express themselves. The evidence of such deficits is rather flimsy—especially in the area of syntax—but so is the counter evidence that there are no deficits, only differences.[3]

Given the present uncertainty over the language abilities of disadvantaged black children, a deficit model might at least have some heuristic value. If there is a deficiency, it seems reasonable to suppose that the most effective intervention will be a direct attack on the root cause of the problem. As indicated above, both vocabulary and syntax are conceivable sources of language deficiency in lower-class children. Since it is not clear whether the disadvantaged black child's putative deficiency stems primarily from one or the other (or from both), a study comparing the effects of training in vocabulary versus sentence construction seems worthwhile. The results might not only provide guidelines for early intervention programs, but also shed new light on the basic controversy over differences and deficits in language development.

METHOD

Subjects

A sample of 72 children, half boys and half girls, was drawn from seven prekindergarten classes at three elementary schools in Richmond, California.[4] At the time of the study, nearly all of the children attending these schools were black, and most of them came from lower-class homes. The total enrollment in the seven classes was approximately 95. Some of these children were eliminated from the study because they were not black, because they were observed to have severe speech impediments, or because they did not attend school regularly. The final sample of 72 was determined by the feasibility of scheduling training sessions which met the design requirements outlined below. Where more than one child was available for a given position, random selections were made. The selected subjects had a mean age of 4½ years when the study began.

Overall Design and Procedure

A pretest battery of language instruments was administered individually to each subject. The children were then divided equally into three treatment groups: vocabulary training, sentence training, and control. The subjects in the two training groups were further divided into groups of four, with two boys and two girls per group when possible. Each of these smaller groups met with the same experimenter twice a week for a 20-minute training session outside of the regular classroom. The training period lasted 6 weeks, during which time the control subjects had virtually no contact with the experimenters. Finally,

---

[3] P. R. Ammon. Restricted vs. elaborated codes and the language of disadvantaged black children. In preparation.

[4] The authors wish to thank the administrators and teachers who provided the necessary subjects and facilities in the Richmond (California) schools.

TABLE 1

ASSIGNMENT OF SUBJECTS IN EACH TREATMENT GROUP TO EXPERIMENTERS A, B, AND C

Phase of study	Treatment group								
	Vocabulary			Sentence			Control		
Pretest	A	B	C	A	B	C	A	B	C
Training	B or C	A or C	A or B	B or C	A or C	A or B	—	—	—
Posttest	A	B	C	A	B	C	A	B	C

all subjects were posttested with the original battery of language instruments.

The experimenters in each phase of the study were three white, college-educated women who had no connection with the regular preschool program. The subjects in each treatment group were assigned to Experimenters A, B, and C according to the design shown in Table 1. (It should be noted that Experimenter C was actually a different person in each phase of the study, due to unforeseen circumstances.) This design balanced the effects of different experimenters on testing and on training, and it permitted a "blind" posttest. As an additional control, each of the regular teachers had approximately equal proportions of children in all three treatment groups. Within all of these constraints, there was random assignment of subjects to testers and to treatment groups.

Materials

Three language instruments were administered as pre- and posttests. The first was Form B of the Peabody Picture Vocabulary Test (PPVT), a widely used recognition test of vocabulary which needs no further description here. The second instrument was a Sentence Imitation Test (SIT) consisting of 50 model sentences which were constructed especially for the present study. In keeping with the focus on syntactic elaboration, these sentences were designed to assess a child's control of several elaborative sentence elements, including prenominal adjectives and possessive nouns, locative prepositional phrases, relative clauses, and adverbial clauses. The experimenter presented each sentence in Standard English with the instruction to "say just what I say." Finally, there was a Picture Interview (PI) patterned after the one used by Loban (1963). In the PI, the experimenter first asked the child some questions about his family, favorite television programs, and so on. Then a series of six pictures was shown and the child was encouraged to talk about each one. All responses to the SIT and to the PI were tape-recorded for later transcription and scoring.[5]

The primary material for vocabulary training was a list of target words drawn from two sources. Thirty-six of the words were taken from Items 20–40 in the PPVT (Form B). Previous data had indicated that these words were unknown by a substantial proportion of children similar to those in the present study. Such blatant "teaching to the test" was a way of establishing whether or not the vocabulary training had any effect at all. Another set of 33 words was suggested by the pictures in the PI. That is, these words could be used to describe or discuss the pictures. The primary material for sentence training consisted of the set of elaborative constructions assessed by the SIT (see preceding paragraph). In teaching these constructions, an attempt was made to use only words which were already familiar to the subjects. A large collection of objects and pictures was developed in order to provide referents for the words and sentence constructions being taught. Each experimenter was equipped with essentially the same set of materials. At no time during training were the test pictures or sentences introduced.

Training Procedures

All three experimenters worked from a common lesson plan for each training session. The methods used for vocabulary training and for sentence training were quite similar. Imitation and group responding were used somewhat to introduce new forms, but the emphasis was on having individual children respond to a variety of referents for each target form. Both recognition and production tasks were employed in the context of gamelike activities, with the ultimate goal being for each child to generalize the production of appropriate words or sentences to new referents. An attempt was made to give every subject an equal number of opportunities to respond. Correct responses were praised and corrective feedback was provided when a form was used incorrectly or when there was no response. Although practically all of the children used some elements of Black English, the sentence training focused not on dialect features, but on the use of elaborative sentence elements which occur in both the standard and the nonstandard dialects of adults. Thus the white experimenters did not emphasize the differences between their own dialects and those of the children.

---

[5] The SIT and PI can be obtained from the authors, along with directions for administration and scoring of both instruments.

Analysis

It was assumed that the random assignment of subjects to treatment groups would preclude significant group differences on the pretest, and the pretest data supported this assumption. Consequently, the effects of training were assessed by comparing group means on five dependent variables from the posttest. A univariate analysis of variance was done for each dependent measure, with two planned comparisons at the .025 level of significance (Hays, 1963). In these analyses, the *n* for each group was 23 or 22, due to incomplete data on some subjects.

RESULTS

Effects of Training on Vocabulary Recognition

The PPVT yields IQ and MA scores, but these normative scores were rendered meaningless by the teaching of test words to subjects in the vocabulary training group. Thus only the mean raw score on the PPVT is shown for each treatment group in Table 2. The difference between the sentence training and control groups was not significant ($F < 1$, $df = 1/66$), but the vocabulary group scored significantly higher than the other two groups combined ($F = 32.78$, $df = 1/66$, $p < .001$). In other words, sentence training had no effect on the recognition of items in the PPVT, but vocabulary training led to an advantage of about 14 words, on the average.

Effects of Training on Vocabulary Production

The PI transcripts were examined for the number of different target words produced by each subject. Various forms of a given word, such as singular and plural, were treated as tokens of a single word type. Not surprisingly, virtually all of the target word types counted in this way came originally from the PI rather than the PPVT. The group means for this measure are shown in the second line of Table 2. There was a significant difference between the vocabulary group and the other two groups combined ($F = 8.17$, $df = 1/65$, $p < .01$), but not between the sentence and control groups ($F < 1$, $df = 1/65$). Thus the results for production of target words parallel the results for recognition.

TABLE 2

GROUP MEANS ON FIVE DEPENDENT MEASURES FROM POSTTEST (23 SUBJECTS PER GROUP)

Dependent measure	E_{MS}	Treatment group means		
		Vocabulary	Sentence	Control
PPVT				
Raw score	93.08	55.3	41.3	41.1
PI				
Target word types	7.46	10.3	8.6[a]	8.0
PI				
Other word types	1419.22	185.7	158.9[a]	170.2
SIT				
Clearly correct	80.11	16.0	14.4	15.7
SIT				
Clearly wrong	63.80	11.5	12.0	13.6

[a] $n = 22$.

Both vocabulary training and sentence training might conceivably have a transfer effect on the production of word types other than those which were selected for training. A count of the other word types produced by each subject in the PI showed no significant difference between the sentence and control groups ($F = 1.02$, $df = 1/65$, $p > .025$) or between the vocabulary group and the other two groups combined ($F = 4.83$, $.025 < p < .05$). Thus only the target words were affected by vocabulary training, and there was no significant evidence that vocabulary or sentence training transferred to the production of other words.

Effects of Training on Sentence Imitation

Two global measures of proficiency in imitation were derived from the SIT. The first, called "clearly correct," was the number of sentences which the subject repeated essentially verbatim, with the allowance of a few very minor deviations. The second score, "clearly wrong," was the number of responses in which the subject obviously failed to preserve the meaning of the model sentence, either by omitting, adding, or substituting content words, or by altering basic grammatical relations, such as subject-object or modifier-head. In between the clearly correct and clearly wrong categories were a number of intermediate deviations, both standard and nonstandard, including changes in tense and number markers, grammatical transformations of the model sentence, and substitution of synonyms. Because many of these responses were ambiguous as to their correctness, they were excluded from the present analysis.

Although the clearly correct and clearly wrong measures are not entirely independent, the difference between them is important in principle because a child could receive low scores on both measures. It was expected that some subjects would translate many model sentences into their own dialects, without getting them clearly wrong. A lot of this translation did occur, but the clearly correct and clearly wrong scores also proved to be highly correlated (the r within treatment groups ranged from $-.82$ to $-.84$ on the posttest). However, the data from both measures are presented in the last two lines of Table 2. Neither variable showed a significant difference between the subjects trained in sentence construction and the other two groups, nor were the contrasts between the vocabulary and control groups significant ($F < 1$, $df = 1/66$, in all cases). It was thought that the effect of sentence training might depend upon the skill of the particular trainer, or upon the pretest performance of the subject on the SIT. Supplementary analyses gave no evidence that either of these factors made any difference. In sum, the SIT showed no direct effects of sentence training, and no transfer effects of vocabulary training.

DISCUSSION

Vocabulary training had a significant positive effect on recognition and production of the target words which were taught. It is interesting to note that, on the PPVT, the mean raw score for the vocabulary group was 55.3. Since a score of 60 represented the top of the range from which target words were selected, retention of the target words appears to have been quite good. As for the production of target words, the superiority of the vocabulary group amounted to a difference of only two words. This may seem like a small gain, but the PI was conducted in such a way that the subject had to produce the target words more or less spontaneously, rather than in response to direct questions. These observations on the extent to which target words were recognized and produced suggest that vocabulary training may have effects which are significant educationally as well as statistically.

Although there was no significant evidence that vocabulary training transferred beyond the specific target words, the present results at least offer some encouragement in this direction. The general effectiveness of vocabulary training might be increased by selecting target words with an eye toward their relationships to each other and to the needs of the child. The words in the present study were selected simply on the basis of the criterion tests, along with some data and intuition about the appropriate level of difficulty.

The positive effects of vocabulary training contrasted markedly with the lack of evidence that sentence training had any effect at all on the measures used in the present study. These findings imply that time devoted to early language training for disadvantaged black children is better spent on vocabulary than on sentence construction. This implication seems most relevant to short-term programs in which a limited amount of time is available for structured language activities. The present results are also consistent with the hypothesis that the syntactic aspects of language development are relatively impervious to direct instruction, perhaps because maturation plays an important role (Cazden, 1968). In this connection, it might even be argued that the present results favor a difference model over a deficit model with regard to the syntactic development of disadvantaged black children. That is, sentence training might have failed to have any effect because the subjects were already elaborating their sentences as much as one might expect for children their age. Conversely, a positive effect of sentence training might have implied the existence of a deficit or gap between readiness and actual achievement in the subjects' control of syntax.

Before the above conclusions are accepted, a number of questions ought to be raised about the negative results obtained with the SIT. The lack of significant group differences probably cannot be attributed to unreliability, since the test-retest correlations for the two SIT measures ranged from .81 to .90 within treatment groups. But it is quite possible to question the validity of the SIT, both as a specific measure of training effects in the present experiment and as a more general test of syntactic capability. First of all, the SIT seems less like the training situation than the PPVT or the PI. That is, the vocabulary subjects practiced pointing to pictures in response to words (as in the PPVT) and they practiced saying words in response to pictures (as in the PI), but the sentence subjects did not practice imitating sentences in the absence of visible referents (as in the SIT). Second, even though there is evidence that a child's imitations are related to the speech he produces and comprehends (Slobin, 1968), the relationship may not be close enough to warrant the use of imitation as a technique for assessing a child's general skill in manipulating syntactic structures. A detailed structural analysis of responses to the PI is in progress in order to provide a further test of training effects, and to check the validity of the SIT as a measure of ability or propensity to produce elaborated sentences.

Finally, it is possible that sentence training might have been more effective with different methods and/or content. Perhaps the target constructions could have been made more salient by means of a communication task which drew even more attention to the elaborated part of each sentence. Also, the target constructions may not have been well chosen. During the training sessions, some constructions appeared to be quite easy for most of the subjects, while others seemed very difficult, perhaps because they were altogether beyond the subjects' grammatical competence. The syntactic analysis of PI responses may indicate that there was a poor match between the training materials and the capabilities of the subjects. In any case, the training and testing of sentence construction skills would appear to be much less straightforward than the training and testing of vocabulary.

REFERENCES

BARATZ, J. C. Teaching reading in an urban Negro school system. In J. C. Baratz & R. W. Shuy (Eds.), *Teaching black children to read.* Washington, D. C.: Center for Applied Linguistics, 1969.

BEREITER, C., & ENGELMANN, S. *Teaching disadvantaged children in the preschool.* Englewood Cliffs, N. J.: Prentice-Hall, 1966.

BERNSTEIN, B. Elaborated and restricted codes: Their social origins and some consequences. In J. J. Gumperz & D. Hymes (Eds.), *The ethnography of communication. American Anthropologist*, 1964, **66**, Pt. 2, 55–69.

CAZDEN, C. B. Environmental assistance to the child's acquisition of grammar. Unpublished doctoral dissertation, Harvard University, 1965.

CAZDEN, C. B. Subcultural differences in child language: An interdisciplinary review. *Merrill-Palmer Quarterly*, 1966, **12**, 185–219.

CAZDEN, C. B. Some implications of research on language development for preschool education. In R. D. Hess & R. M. Bear (Eds.), *Early education.* Chicago: Aldine, 1968.

DEUTSCH, M. The disadvantaged child and the learning process. In A. H. Passow (Ed.), *Education in depressed areas.* New York: Teachers College, Columbia University, 1963.

DICKIE, J. P. Effectiveness of structured and unstructured (traditional) methods of language training. In M. A. Brottman (Ed.), Language remediation for the disadvantaged preschool child. *Monographs of the Society for Research in Child Development*, 1968, **33** (8, Whole No. 124).

HAYS, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

LABOV, W. The logic of nonstandard English. In F. Williams (Ed.), *Language and poverty.* Chicago: Markham, 1970.

LOBAN, W. *The language of elementary school children.* (Research Report No. 1). Champaign, Ill.: National Council of Teachers of English, 1963.

OSBORN, J. Teaching a teaching language to disadvantaged children. In M. A. Brottman (Ed.), Language remediation for the disadvantaged preschool child. *Monographs of the Society for Research in Child Development*, 1968, **33** (8, Whole No. 124).

RAPH, J. B. Language development in socially disadvantaged children. *Review of Educational Research*, 1965, **35**, 389–400.

SLOBIN, D. I. Imitation and grammatical development in children. In N. S. Endler, L. R. Boulter, & H. Osser (Eds.), *Contemporary issues in developmental psychology.* New York: Holt, Rinehart & Winston, 1968.

Journal of Educational Psychology
1971, Vol. 62, No. 5, 427–431

INTERACTIVE EFFECTS OF ACHIEVEMENT ORIENTATION AND TEACHING STYLE ON ACADEMIC ACHIEVEMENT[1]

GEORGE DOMINO[2]
Fordham University

This study tests the hypothesis that there is an interaction between a student's achievement orientation and the teaching style he is exposed to, which differentially affects both the amount of learning that takes place and the degree of expressed satisfaction with the scholastic environment. One hundred students, selected because of their extreme scores on the Achievement-via-Conformance and Achievement-via-Independence scales of the CPI, were assigned to introductory psychology sections taught in either a conforming or an independent manner. An analysis of their scores on a final examination consisting of multiple-choice items and essay questions, as well as their ratings of teacher effectiveness and course evaluation, indicated a clear interaction effect. Students taught in a manner consonant with their achievement orientation obtained significantly higher means on the multiple-choice items, on factual knowledge ratings of their essay answers, and on their ratings of teacher effectiveness and course evaluation, than their peers taught in a dissonant manner.

Although both educators and researchers are concerned with maximizing academic achievement, little empirical work has been reported that attempts to systematically alter selected aspects of the college environment in an effort to directly influence academic achievement.

Domino (1968) studied the academic performance of high and low scorers on the Achievement-via-Conformance (Ac) and Achievement-via-Independence (Ai) scales of the California Psychological Inventory (Gough, 1957) to test the hypothesis that conforming and independent achievement orientations are differentially related to academic achievement. The results not only supported this notion, but indicated that for students high on one achievement dimension and low on the other, there was a distinct interaction between the student's achievement orientation and the demands of the college environment. Students high on Ac obtained higher grades in courses rewarding conforming behavior whereas students high on Ai obtained higher grades in independence rewarding courses.

The present study is concerned with the general hypothesis that there is an interaction between a student's achievement orientation and the teaching style he is exposed to, and that this interaction will differentially affect both the amount of learning that takes place and the student's expressed satisfaction with his scholastic environment. Specifically, it is hypothesized that students high on Ac or Ai who are taught in a manner consonant with their achievement orientation will perform better academically and report greater satisfaction than their peers who are taught in a manner dissonant with their achievement orientation.

METHOD

Sample

In September 1968, the entering class (N = approximately 900) of a large university was administered the CPI as part of a freshman testing program. In September 1969, from a frequency distribution of scores on the Ac and Ai scales, the top 50 high Ac-low Ai and the top 50 low Ac-high Ai students were identified, eliminating those whose SAT scores fell below 550 or above 650. These 100 students were assigned to four introductory psychology sections, with the same instructor, in such a manner that each section contained 25 students, with equal sex composition, comparable mean SAT scores, but homogeneous

---

[1] This research was made possible by a research fellowship from American College Testing Program. The author is grateful to the staff of ACT, and in particular to Nancy Cole, for their assistance.

[2] Requests for reprints should be sent to George Domino, Department of Psychology, Fordham University, Bronx, New York 10458.

TABLE 1

SUMMARY STATISTICS FOR FOUR SECTIONS

Variables	Sections			
	High Ai	High Ai	High Ac	High Ac
CPI Ac				
range	15–25	16–25	28–32	28–32
M	20.7	21.2	29.4	29.2
SD	2.9	2.2	1.6	1.5
CPI Ai				
range	22–26	21–26	12–19	14–19
M	23.7	23.5	17.4	17.3
SD	1.4	1.4	1.5	1.2
Sex composition				
Male	11	9	12	12
Female	13	12	13	11
Total	24	21	25	23
SAT Verbal				
M	612.5	612.2	595.4	604.6
SD	32.2	26.8	26.2	32.5
SAT Math				
M	602.3	594.5	588.1	584.6
SD	35.8	26.5	34.6	32.7

Note.—Twelve males and 13 females were assigned to each section. Withdrawals reduced the registration to the numbers shown under sex composition. All values are for the participating students.

with respect to their Ac or Ai scores. Table 1 presents a demographic description of the four sections.

Procedure

Neither the instructor nor the students were aware that class assignments had been made on other than routine scheduling by the registrar's office. The instructor, however, cooperated with the author in a study of the effectiveness of teaching styles, by teaching Sections 2 and 3 in a "conforming" manner, and Sections 1 and 4 in an "independent" manner, using the criteria indicated by Domino (1968). Thus for one group of high Ai students and one group of high Ac students, instructional material was presented solely through lectures, great emphasis was placed upon factual knowledge, classroom attendance was required, and course content closely paralleled textbook assignments. For the high Ai and high Ac sections taught in an independent manner, emphasis was placed upon ideas rather than facts and upon the active participation of students in the learning process. All four sections were assigned identical textbook readings and examinations, but for the independent sections the examinations served as a guide for the student rather than a grading device.

Instruments

At the termination of the semester all students were administered a final examination consisting of 200 multiple-choice items tapping mostly factual content, as well as six essay questions which demanded both convergent and divergent thinking (Guilford, 1959). Both the multiple-choice items and the essay questions were formulated prior to the beginning of the course and independently of the course instructor.

Essay answers for each student were independently rated by three psychologists using two 9-point scales, one scale rating amount of factual knowledge present and the other degree of original thinking present. All judges rated all essay answers on both dimensions, with order of ratings randomized for both dimensions.

Students evaluated both the course and the instructor using 7-point rating scales; although code numbers were used to provide anonymity and to obtain honest responses, it was possible to compare these ratings with other variables. Finally, for each student, his introductory psychology grade (assigned without reference to the final examination) and his cumulative GPA for the first 2 years (not considering his introductory psychology grade) were available.

The data were submitted to both a correlational analysis and a two-factor analysis of variance.

RESULTS

Table 2 presents a correlation matrix for the 11 variables of concern in this study. Of the 55 correlation coefficients, 16 reach statistical significance, 11 at the .01 level, and 5 at the .05 level.

A preliminary question needs to be asked regarding the interrater reliability of the factual knowledge and original thinking ratings of essay answers. As can be seen in Table 3, the interrater reliability coefficients are in an acceptable range, although not exceptionally high. In subsequent analyses, ratings were pooled using either the mode if at least two judges agreed, or the median if no agreement was present. Table 3 also indicates the correlations between each rater and the pooled values.

A second concern is in relation to the independence of both the factual knowledge-original thinking ratings and the teacher effectiveness-course evaluation ratings. Factual knowledge and original thinking ratings did not correlate significantly with each other, as indicated in Table 2. Factual knowledge ratings do correlate significantly with scores on the multiple-choice items, with introductory psychology grades, and with GPA; original thinking ratings also correlate significantly with introductory psychology grades, as well as with CPI Ai and,

TABLE 2

CORRELATIONS BETWEEN INDICATED VARIABLES FOR ALL STUDENTS

	2	3	4	5	6	7	8	9	10	11
CPI Ac[a]	−83**	−12	−22	−02	−11	−60**	−21	−02	−30**	−15
CPI Ai[a]		12	22	02	11	60**	21	02	30**	15
SAT Verbal			22	−12	02	18	−11	−07	−06	01
SAT Math				−02	01	05	03	−19	01	05
Multiple choice					50**	21	21	25*	55**	38**
Factual knowledge						16	16	22	44**	25*
Original thinking							25*	10	24*	22
Teacher effectiveness								27**	39**	03
Course evaluation									18	14
Psychology grade										24*
GPA										

Note.—$N = 93$.

[a] Correlation coefficients between Ac, Ai scores and other variables are biserial; since extreme scoring subjects were selected, the obtained *r*s may be slightly inflated.

* $p < .05$.

** $p < .01$.

Ac scores, and with students' ratings of teacher effectiveness. These correlations support both the independence of the factual knowledge and original thinking ratings, and their construct validity.

Student ratings of teacher effectiveness and course evaluation do correlate significantly with one another, but in a rather moderate manner. Ratings of teacher effectiveness correlate significantly with original thinking ratings and with introductory psychology grades. Ratings of course evaluation correlate significantly with scores on the multiple-choice items. In general, these results also support the construct validity of the two effectiveness rating scales.

The relevancy of the multiple-choice items can also be questioned. It would be impossible, and basically not relevant, to argue that the items represent the basic knowledge a student ought to acquire in an introductory psychology course. What can be argued is that the items are in fact tapping cognitive processes of relevance to traditional academic achievement; scores on the multiple-choice items did indeed correlate significantly with both psychology grades and with GPA.

Table 4 presents the results of the analyses of variance performed on each of the seven dependent variables. Since selection limits were placed on SAT scores, and since SAT scores did not correlate significantly with other variables, analysis of covariance was not deemed necessary. For the multiple-choice items, factual knowledge ratings, teacher effectiveness ratings, course evaluation ratings, and introductory psychology grades, there are significant interaction effects between achievement orientation and teaching style. For the original thinking ratings, teacher effectiveness ratings, introductory psychology grades, and GPA, there are significant achievement orientation effects. No significant teaching style main effects were obtained.

DISCUSSION

The major concern of this study was with the hypothesized interaction between student achievement orientation and teaching style. Despite the artificiality of having the same instructor role play both independent and conforming teaching styles, the results quite clearly indicate a very definite interaction between student achievement orientation and teaching style. If it can be argued that educational aims should include both the imparting of factual knowledge and some degree of satisfaction on the part of students, then one method of achieving these aims is to match student achievement orientation with teaching style. On the other hand, the results indicate that teaching style has no effect on original thinking; to elicit original

TABLE 3

INTERRATER RELIABILITY COEFFICIENTS

Factual knowledge				Original thinking			
Rater	B	C	Pooled	Rater	B	C	Pooled
A	67	77	91	A	55	75	88
B		58	79	B		55	73
C			88	C			90

TABLE 4

SUMMARY STATISTICS AND ANALYSES OF VARIANCE RESULTS ON DEPENDENT VARIABLES

Variable	Group[a]	*M*	*SD*	Results of ANOVA	
				Source	*F*
Multiple choice	1	98.2	21.6	Achievement orientation	0.00
	2	89.6	25.1	Teaching style	1.08
	3	103.5	29.6	Interaction	6.50*
	4	84.7	23.7		
Factual knowledge ratings	1	6.7	1.7	Achievement orientation	.75
	2	5.7	2.0	Teaching style	.83
	3	6.7	1.7	Interaction	10.68**
	4	5.0	2.2		
Original think ratings	1	7.0	1.3	Achievement orientation	66.73**
	2	7.1	1.2	Teaching style	.19
	3	4.2	1.5	Interaction	.33
	4	4.6	2.1		
Teacher effect	1	6.8	.5	Achievement orientation	7.66**
	2	6.3	1.1	Teaching style	2.31
	3	6.5	.8	Interaction	14.42**
	4	5.5	1.1		
Course evaluation	1	6.7	.6	Achievement orientation	.54
	2	5.5	1.7	Teaching style	.47
	3	6.3	1.2	Interaction	12.07**
	4	5.6	1.4		
Psychology grade	1	3.5	.5	Achievement orientation	8.66**
	2	2.9	.8	Teaching style	.68
	3	3.1	.8	Interaction	20.54
	4	2.3	.8		
GPA	1	2.8	.5	Achievement orientation	6.31*
	2	2.7	.5	Teaching style	.24
	3	2.5	.4	Interaction	.02
	4	2.6	.4		

[a] Group 1 = high Ai with independent teaching style; Group 2 = high Ai with conforming teaching style; Group 3 = high Ac with conforming teaching style; Group 4 = high Ac with independent teaching style.

* $p < .05$.

** $p < .01$.

thinking from students one must begin with students whose achievement orientation is conducive to independent, original thinking. These results can be seen as either disconcerting or reassuring depending upon what one believes are the effects of college courses on students' original thinking abilities.

The obtained significant interactions have implications for educational practices which can be perceived in rather different ways. One may argue, for example, that homogeneous grouping of students on the basis of their achievement orientation is a desirable practice, since it maximizes both their factual learning and their reported satisfaction. Yet, homogeneous grouping may deny both the independent and the conforming student the valuable insights that result from exposure to heterogeneity. It is quite clear, however, that the interaction of achievement orientation with teaching style is an important dimension which has been seriously neglected by both educators and researchers. It is possible to provide students with the type of educational setting that would most effectively utilize their potential, and also to expose students to diversity of opinions and achievement orientations. Only by first recognizing these dimensions, however, can their relevancy to self-actualization be utilized in a positive manner.

Significant achievement orientation differences were obtained on original thinking ratings, teacher effectiveness ratings, introductory psychology grades, and GPA with the high Ai students scoring higher. For both the original thinking ratings and GPA there are no significant interaction effects; students high on Ai exhibit more original thinking and perform better academically than their Ac peers. The latter result agrees with the findings of Domino (1968) for students in a different college. For the teacher effectiveness ratings and the introductory psychology grades there are both achievement orientation differences as well as interaction differences.

Finally, there were no differences attributable solely to teaching style. It is understandable that a teacher does not operate in a vacuum, despite the vacuous expressions of students in Monday morning classes; it is crucial, however, to discover what variables besides achievement orientation are of interactive significance.

REFERENCES

DOMINO, G. Differential prediction of academic achievement in conforming and independent settings. *Journal of Educational Psychology*, 1968, 59, 256–260.

GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto, California: Consulting Psychologists Press, 1957.

GUILFORD, J. P. Three faces of intellect. *American Psychologist*, 1959, 14, 469–479.

(Received August 3, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 5, 432–438

EFFECTS OF VERBAL WARNING UPON MISAPPLICATION OF A RULE OF LIMITED APPLICABILITY[1]

DUNCAN MCQUARRIE[2] AND ARDEN GROTELUESCHEN[3]
Center for Instructional Research and Curriculum Evaluation
University of Illinois

The effects of level of intensity of a verbal warning upon the tendency to misapply a rule of limited applicability were investigated using a fixed effects two-factor multiple dependent variable experimental design. The subjects consisted of 86 undergraduate teacher education students. Results showed the introduction of the rule of thumb to have a detrimental effect on transfer task performance. The three levels of intensity of the warning resulted in differential performance of the experimental groups on three dependent variables. The relationship between the effectiveness of the warning statements and the level of original learning of the more general principle of the calculation of significant figures was discussed.

A recent study (Logan & Wodtke, 1968) offers experimental evidence relevant to a frequent strategy of classroom teachers, particularly science and mathematics teachers. This strategy involves the teaching of a general principle (applicable to a wide variety of transfer problems) followed by the teaching of a "rule of thumb" (applicable to a narrow class of these same transfer problems).[4]

These rules of thumb typically are simple, and hence are justified as shortcut, time-saving procedures. Logan and Wodtke (1968) suggest that because of the ease of use, students have a strong set to apply a rule of thumb regardless of its restricted applicability. For this reason, rules of thumb are potential sources of negative transfer having a high probability of occurrence. If students are not exposed to the rule of thumb limitation, fail to learn the limitation, or are unable to recall the limitation during the transfer task, they will have a strong tendency to overgeneralize its use.

The findings of Logan and Wodtke (1968) generally support the above speculation concerning the effects of teaching a rule of thumb. Subjects taught a rule of thumb as part of an instructional program designed to teach a more general principle did significantly poorer on a transfer task than subjects taught only the general principle. They attribute this poorer performance to the misapplication (negative transfer) of the rule of thumb. This misapplication occurred even though the rule of thumb instructions included a didactic warning about the limitation of the rule. It was assumed that the warning was ineffective and therefore, subjects were unable to determine the appropriateness of the rule of thumb during the transfer task. If this interpretation is correct, any procedure that would facilitate the original learning and/or subsequent recall of the limitations of the rule of thumb should reduce the cases of misapplication. It is quite likely that modifications (e.g., discrimination training, requiring subjects to make overt responses relevant to the limitation) to the program used by Logan and Wodtke (1968) would have resulted in fewer misapplications of the rule of thumb. However, the objective of the present investigation was not to demonstrate how the program per se might be improved (although that would certainly be a worthwhile endeavor). Instead it was desired to further clarify the effect of using a verbal warning.

Logan and Wodtke (1968) point to the possibility that the verbal warning employed was "not strong enough to call their (subjects) attention to the limitations of the rule." The warning used was selected a priori after considering the relative intensity of various alternative forms. It was concluded that verbal warnings alone may not be sufficient to keep subjects from misapplying the rule of thumb.

---

[1] The authors acknowledge the assistance of Herman L. Glaess for data collection, Tracy H. Logan for the original study materials, and James L. Wardrop for statistical analysis.

[2] Now at Central Washington State College.

[3] Requests for reprints should be sent to Arden Grotelueschen, College of Education, Education Building, University of Illinois, Urbana, Illinois 61801.

[4] See the article by Logan and Wodtke (1968) for a more definitive distinction between principle and rule.

To diminish the role of a warning would seem to be counter to what goes on in everyday practice. It is common for speakers and writers to increase the stimulus intensity of important messages. Speakers raise their voices in emphasis of a point. Writers underline key phrases and words. In the research being reviewed this was the role of the verbal warning, to emphasize the importance of remembering the limitations of the rule of thumb. The implicit assumption was that the stronger the warning, the more likely the limitations would be recalled during the transfer task.

The aim of this study was to investigate the effects of the intensity of a warning upon the frequency of misapplications of a rule of limited applicability. Specifically, the effects of three intensities of a verbal didactic warning (experimental conditions) upon the tendency to misapply a rule of thumb were studied. It was hypothesized that if intensity of the warning is increased, then the occasions of misapplication of the rule will decrease. As a further clarification of the interpretation presented by Logan and Wodtke (1968) a more direct check on the subject's ability to recall the didactic warning was also made.

METHOD

Subjects

The subjects for this investigation consisted of 86 undergraduates from four education classes (two upper-level courses and two lower-level courses) at Concordia Teachers College, Seward, Nebraska. The experiment was conducted in the spring of 1969 during the regular evening class period for these classes. All subjects were enrolled in the teacher education program. Each subject was paid two dollars for participating in the study.

Design

The experimental design was a fixed effects 4 × 2 factorial with multiple dependent variables. Because the subjects came from existing classes, they were randomly assigned within classes to the four levels of verbal warning (three experimental and one control) and two classifications of test. One test was a measure of original learning and the other was a measure of dogmatism (Rokeach, 1960) used to control for time as well as to provide a measure of personality considered to be of theoretical interest.

Materials

The instructional materials for this investigation were an adaptation of the branching self-instructional program described by Logan and Wodtke (1968). This program is designed to teach the topic of significant figures at the level of high school or introductory college physics. The programmed booklet consists of two sections. The first section was identical (with a few slight editorial and format modifications) to that described by Logan and Wodtke (1968) as "Version 1." It consists of maintrunk frames, remedial frames, and several practice problems designed to teach the general principle of significant figures. The length of this section was 117 pages. All of the instruction was provided within the context of multiplication problems.

The second section material varied depending upon the treatment condition. For the experimental conditions the program was designed to teach subjects a rule of thumb of significant figures. This rule has limited applicability in that it only applies to multiplication and division problems. Incorrect solutions are obtained when the rule is used with addition and subtraction operations. This material parallels in content that added by Logan and Wodtke (1968) to form their "Version 2." The first page of this 12-page section introduces the material as "A quicker way for determining which figures in a calculation are significant." This is followed by a paragraph that includes a verbal warning about the limitations of the rule. Immediately after the warning appears the rule, followed by examples and practice problems. For each experimental condition the strength of the warning was varied to form the treatment. (Examples of the actual warnings as well as a description of the method used to derive the scale values for their respective strengths are given below.)

The second section material for the control condition consisted of 11 pages of material on the history of standardized measurement adapted from a similar program used by Grotelueschen (1967).

Warning Factor

Using the warning employed by Logan and Wodtke (1968) as a model, 27 additional parallel warnings were constructed varying with respect to their relative intensity. These 28 statements were scaled using a limited category/multiple stimulus procedure (Torgerson, 1958). Using a model for equal appearing intervals scaling provided by Tucker (1967) these data were analyzed to give estimates of scale value and standard deviation for each warning statement.

TABLE 1

WARNING STATEMENTS OF EXPERIMENTAL CONDITIONS WITH SCALE VALUES AND STANDARD DEVIATIONS

Statement	Scale value	*SD*
Strong warning: "It is absolutely necessary that you remember that it is just a rule of thumb, and works only for products and quotients, not for addition."	10.42	1.75
Medium warning: "It is just a rule of thumb. and works only for products and quotients, not for addition [Logan and Wodtke, 1968]."	5.16	1.97
Mild warning: "This rule of thumb is a short-cut method to be used with multiplication and division operations."	.30	1.94

It was desired to have statements representing each of the extremes and one representing a middle position. The mild warning selected had a scale value of .30 while the strong warning had a scale value of 10.42. Because the scale value of the warning statement used by Logan anff Wodtke (1968) conveniently turned out to be in the middle range of intensity (strength = 5.16) it was selected as the medium value of strength or warning. These three warnings, their scale values, and standard deviations are given in Table 1.

Test Factor

Since level of original learning is an obvious source of confounding in an investigation of transfer or interference, a measure of the level of original learning was planned. This test consisted of eight completion items—five multiplication and three division—designed to assess the information presented in the first section of the instructional material. In order that the possible practice effects of taking such an original learning test could be assessed only half the subjects were administered it. The other half of the subjects were given an attitude scale designed to give one index of their authoritarianism. This measure was a 40-item dogmatism scale (Rokeach, 1960). This instrument was chosen because it was substantively unrelated to the learning material and because it required a similar amount of time to complete as the original learning test. In addition, this scale was thought to provide a measure of a construct theorized to have important relationships to the effect of warnings. For example, it was expected that there would be a positive correlation between the dogmatism score and the likelihood of attending to a didactic warning.

Dependent Variables

Three dependent variables were considered to be of importance to this study. One variable consisted of the subject's performance on a posttest designed to measure application of a general principle for computing significant figures of arithmetic operations. This test was composed of 13 completion items: two multiplication, two division, five addition, and four subtraction items. The transfer performance score was defined as the number of correct division, addition, and subtraction items.

A second variable was the subject's ability to recall the operations for which the rule of thumb was appropriate (multiplication and division) and those operations for which the rule of thumb was inappropriate (addition and subtraction). A subject's score on this variable was defined as the number of operations correctly identified. This measure was collected as a part of a postexperiment questionnaire designed to collect attitudinal and demographic information about the subjects.

The third variable was a measure of the amount of negative transfer (inappropriate use of the rule) exhibited on the posttest. A subject's score on this variable was defined as the number of errors on the transfer items that could be attributed to the misapplication of the rule of thumb. An error was attributed to the misapplication of the rule if the subject gave the answer that would have resulted from the use of the rule and if there was no evidence (i.e., scratch work) suggesting another strategy had been used. These three dependent variables have been labeled Transfer, Recall, and Misuse, respectively.

Procedure

All subjects attended a group session held in a large tiered classroom. Each subject was seated together with subjects receiving the self-instructional booklets containing the same Warning condition. All subjects had previously been randomly assigned to the Warning condition within the class in which they were enrolled. The subjects receiving similar conditions were randomly assigned an experimenter who provided directions and administered the project materials during the course of the experiment. After the project director presented initial instructions all subjects were asked to open their booklets and to complete the first section (117 pages) of the booklet. They were also asked to record the time that they began and completed this section. Upon completing the first section subjects were directed by information in the booklet to signal their experimenter who then randomly assigned pairs of subjects, as they finished the first section, to either the original learning or the dogmatism Test condition. This procedure equated the completion rate of subjects in each of the classifications of Test. After the subject completed his original learning or dogmatism test he was asked to open the instruction booklet to page 118 and begin the second section. When the subject completed this section he again signaled his experimenter who then collected the completed booklet and administered the posttest. After completing the posttest the subject was administered a postexperiment questionnaire. Upon completing the questionnaire he was paid $2 and asked to remain seated until all subjects completed their tasks.

RESULTS

It was hypothesized that increasing the intensity of a warning condition would have an effect upon several subsequent preformance variables. Specifically it was anticipated that as intensity increased: (*a*) the occasions of misapplication of the rule would decrease; (*b*) the ability to recall the rule (limitations) would increase; and (*c*) the transfer score would increase. It was further hypothesized that there would be no difference on the three variables between the Test conditions.

While the three dependent variables are logically related, they do represent subtle facets of the possible effects of the Warning condition. For this reason, a multivariate analysis of variance (MANOVA) was used to simultaneously analyze the three dependent variables.

The raw scores on the dependent variables are properly conceived as proportions of a "total possible score." Therefore, raw scores were converted to proportions. An arc sine transformation was used to convert these scores on the three dependent variables to angles expressed in radians. Such a transformation has been shown to effectively eliminate the undesirable dependency (Myers, 1966) between the population mean and standard deviation of proportional data.

A two-way MANOVA following the procedures outlined by Bock and Haggard (1968) was used to analyze the transformed data. The results indicate that there is a multivariate effect attributable to the Warning condition ($F = 2.66, p < .006$), while the multivariate difference for the Test condition and the interaction between the Warning and Test conditions can be accounted for by chance alone (Fs = .62 and 1.03; ps < .61 and .42, respectively).

TABLE 2

UNIVARIATE AND STEP-DOWN ANALYSES OF THE WARNING CONDITION

Variable	Univariate F^a	p	Step-down F	p
Misuse	3.88	<.01	3.88	<.01
Recall	3.87	<.01	3.72	<.02
Transfer	2.48	<.07	.56	<.65

[a] For hypothesis, $df = 3$; for error, $df = 77$.

TABLE 3

RAW AND TRANSFORMED (ARC SINE) MEANS BY LEVEL OF WARNING CONDITION AND DEPENDENT VARIABLE

Warning condition	n	Misuse		Recall		Transfer	
		Raw	Arc sine	Raw	Arc sine	Raw	Arc sine
No rule	22	1.18	.27	1.50	.61	5.86	.81
Strong	21	3.86	.68	2.86	1.12	4.05	.63
Medium	20	3.85	.69	2.50	.98	3.85	.61
Mild	22	4.14	.73	1.41	.57	3.41	.51

Note.—Arc sine values of proportions are expressed as radians.

Table 2 shows a more detailed analysis of the Warning effect. The univariate Fs suggest that the groups differ on each of the three dependent variables. (The means for the three dependent variables of the raw and transformed scores for each level of the Warning condition are presented in Table 3.) However, the step-down Fs suggest that while the groups differ with respect to the Misuse ($p < .01$) and Recall ($p < .02$) variables, the difference on the Transfer variable can be accounted for by the Misuse and Recall variables.

The significant multivariate F for the Warning condition suggested that a nontrivial model could be fit to the data. The additive main-effects model was assumed and estimates of linear functions of the effects of that model were derived using the method of least-squares. The linear functions chosen were the contrasts between the means of the Medium Warning group and the means of the remaining three Warning groups. These simple contrasts were chosen for two reasons. The contrast of the Medium group and the No Rule (control) group corresponded to the comparison used by Logan and Wodtke (1968) and thereby helped to relate the two studies. Contrasting each of the Strong and Mild groups with the Medium group was a logical way to analyze the differences among the experimental conditions.

A formal equivalent of discriminant analysis was used to help portray the differences among the Warning groups. Since the step-down test suggested that the Misuse and Recall variables would account for the variance in the Transfer scores, the Transfer variable was not entered into the discriminant anal-

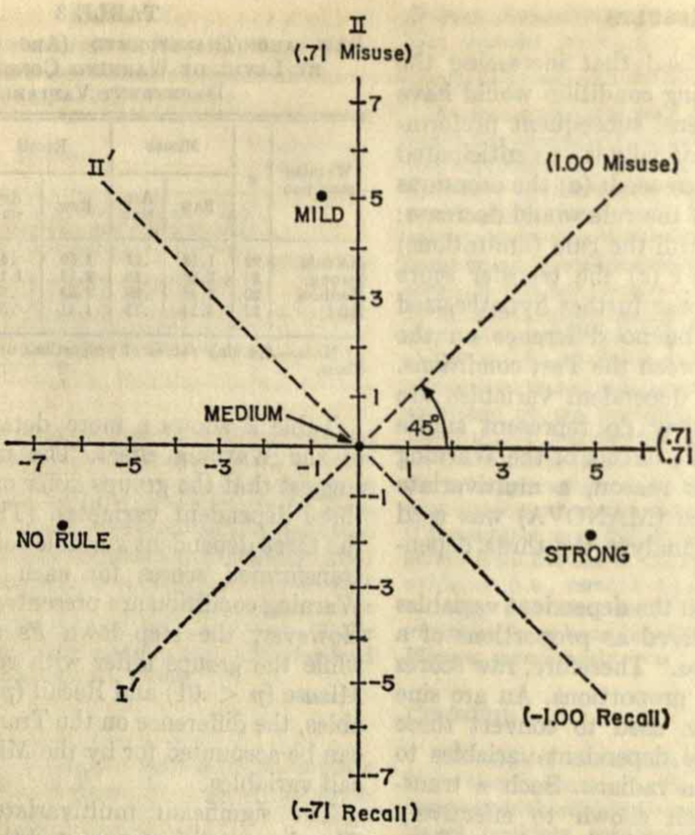


FIG. 1. Plot of four Warning groups on two discriminant axes (I and II), each axis defined by the standardized weights of the variables in the corresponding discriminant functions. (Rotated (45°) axes (I′ and II″) are represented by broken lines.)

ysis. For the Misuse and Recall data the two discriminant functions accounted for 71.50% and 28.44% of the canonical variation. The discriminant functions for the Warning effects are:

$$v_{\text{Warning}},\ \text{I} = 1.378\ \text{Misuse} + 1.098\ \text{Recall},$$
$$v_{\text{Warning}},\ \text{II} = 1.375\ \text{Misuse} - 1.102\ \text{Recall}.$$

Bartlett's chi-square test for the significance of successive discriminant functions yielded $\chi^2 = 21.40$, $p < .002$ for the two discriminant functions and $\chi^2 = 6.34$, $p < .04$ for the residual after removing the first function. These results suggest that the Warning groups differed on two independent dimensions. By applying these two discriminant functions to the least-squares estimates of the effects associated with the contrasts, a canonical representation of the contrasts was obtained and has been displayed graphically in Figure 1. Since the Medium group has been contrasted with the other three Warning groups it is implicit in the origin of the coordinate system.

In order to interpret the discriminant functions, the coefficients were converted to standard form. These results show that the Misuse and Recall variables make an equal contribution to the group differences along with the first discriminant axis (I) (standardized weight of .71 for each variable), while for the second axis (II), the standardized weights for the Misuse and Recall variables are of equal magnitude but opposite in sign (.71 and −.71, respectively).

Having shown the groups to differ in a two-dimensional space it was desired to represent the reference axes of that space in a more parsimonious manner. A rigid rotation of 45° [5] (See Figure 1) resulted in axes defined by single variables. The first rotated axis (I′) was defined by Misuse (1.00) while the second rotated axis (II′) was defined by Recall (−1.00).

DISCUSSION

The warning statements selected on the basis of the scaling procedure yielded differences which were reflected in a two-dimensional space. The contrast exhibiting the strongest effect was between the No Rule and Medium groups, i.e., between the control and middle intensity conditions. Considering this contrast to be representative of the difference between the control condition and the experimental conditions, the effect is characterized by a combination of the Misuse and Recall dimensions with the former having a slightly greater contribution. The two contrasts representing the differences among the experimental conditions show effects of approximately equal magnitude, but differing in orientation. For both the contrast between the strong and the Medium groups and the contrast between the Mild and the Medium groups the contribution of the Misuse dimension was moderate and in the same direction. The Recall dimension separated the experimental treatment conditions equally, but in opposite directions from the Medium group.

The symmetry of these effects seems to reflect the equal spacing of the intensity of the warning statements. It is important to note that even though the Mild and Strong Warning conditions represented the extreme intensities of the scaled statements, there was a small overlap of the distribution values for each of these conditions with the Medium group. If more extreme statements had been constructed, so that the overlap with the Medium group would have been eliminated, even greater effects due to the intensity of warning may have been obtained.

The projections of the Mild and No rule groups on the second rotated axis showed them to differ almost equally from the Medium group. If the performance of a No Rule group is assumed to represent a level of chance performance on the Recall variable, then it could be concluded that the ability to recall the substance of a warning of mild intensity is not likely to be above chance performance either. The use of a control group which would have received the rule of thumb without a warning may have provided a better control group for determining the effect of a warning.

The discriminant analysis results discussed above suggest that there are differences between the control and the experimental groups for the Warning condition. These findings are supported by the obtained mean differences between the control and the experimental groups on each of the three dependent variables. Thus, these results confirm those obtained by Logan and Wodtke (1968).

In the present study the effects of the intensity of warning among the three experimental groups was of primary interest. For the rotated two-dimensional space of the discriminant analysis the Misuse dimension (I′) indicated no clear differentiation among the experimental groups, but the Recall dimension (II′) did differentiate among the experimental groups. Furthermore, the means of the experimental groups for all three dependent variables were generally ordered in the predicted direction with the strongest effect observed for the recall variable. The consistency of these results warrant the conclusion that the experimental Warning condition had a pronounced effect.

---

[5] Maxwell (1961) has pointed out that the original discriminant functions are not the only sets of weights that will produce the group differences in the discriminant space. Any rigid rotation of the orthogonal axis will produce new sets of weights that are jointly equivalent to the original discriminant weights but which might have meaningfulness in their own right. Anderson, Walberg, and Welch (1969) used multiple discriminant functions with subsequent rotation of discriminant axes as an analysis procedure in a nonorthogonal design. It was claimed that it is necessary to calculate the "discriminant loadings," which are defined as "roughly analogous to factor loadings in factor analysis," in order to rotate discriminant functions. The justification for this stipulation is unclear and we could see no advantage in following their procedures. We, instead, chose to use the discriminant weights in performing the rigid rotation and followed the procedures for deriving the transformed weights given by Harman's (1967, p. 300) Formula 14.13.

It was expected that the Recall and Misuse variables would be inversely related. They were, however, observed to be uncorrelated ($r = .02$) for the experimental groups. This finding was unexpected because ability to recall the rule limitation would seem to be a necessary condition for avoiding misapplication of the rule. However, because subjects were asked to recall the rule limitation as part of a postexperiment procedure, it is possible that the Recall posttest elicited a response that may not have been operable during the earlier Transfer test.

Additional evidence from the present study suggests an alternative interpretation. The mean performance on the five-item completion-type original learning test (similar to test events in the instructional material), administered to subjects as part of the Test factor, was only 1.9. If subjects have not mastered the general principle in the instructional material they may be more likely to misapply the rule of limited generality. In the present study, weak support for this was found in the small negative correlation of $-.15$ between original learning test scores and number of misapplications. If a warning accompanying a rule is conceptualized as a set of directions about when to apply a general principle or when to apply a rule of limited generality, the expectation is that the warning can only be functional if both principle and rule have been mastered by the subject. To the extent that the general principle was not mastered in the present investigation the subjects did not have a real choice between applying the general principle or the rule. Under an assumption that subjects had no choice, the warning accompanying a rule would not be expected to have the desired effect.

Given this interpretation, it is likely that level of original learning of the general principle attenuated the correlation between the Misuse and Recall variables. That is, even though subjects might have been able to recall the rule limitations, they may not have been able to apply the more difficult general principle, and, hence, applied the simpler rule. Additional research, designed to manipulate both the intensity of the warning and level of mastery of instruction, is necessary to verify this interpretation.

The results from the MANOVA have provided a better understanding of the effects of using a warning as an instructional strategy. In one sense they have verified the findings of previous research and the hypotheses of the present investigation. Further research should be designed to demonstrate the generalizability of these effects by using different subject matters or problem-solving situations.

REFERENCES

ANDERSON, G. J., WALBERG, H. J., & WELCH, W. W. Curriculum effects on the social climate of learning: A new representation of discriminant functions. *American Educational Research Journal*, 1969, 6, 315–328.

BOCK, R. D., & HAGGARD, E. A. The use of multivariate analysis of variance in behavioral research. In D. K. Whitla (Ed.), *Handbook of Measurement and Assessment in Behavioral Sciences.* Reading, Mass.: Addison-Wesley, 1968.

GROTELUESCHEN, A. *Differentially structured introductory learning materials and learning tasks.* (Final Report on United States Office of Education Grant No. 1-7-002447-2042) Washington, D.C.: United States Government Printing Office, 1967.

HARMAN, H. H. *Modern factor analysis* (2nd ed.). Chicago: University of Chicago Press, 1967.

LOGAN, T. H., & WODTKE, K. H. Effects of rules of thumb on transfer of training. *Journal of Educational Psychology*, 1968, 59, 147–153.

MAXWELL, A. E. Canonical variate analysis when the variables are dichotomous. *Educational and Psychological Measurement*, 1961, 21, 259–271.

MYERS, J. L. *Fundamentals of experimental design.* Boston: Allyn and Bacon, 1966.

ROKEACH, M. *The open and closed mind.* New York: Basic Books, 1960.

TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

TUCKER, L. R. Model for equal appearing intervals scaling. (Unpublished mimeographed material) Department of Psychology, University of Illinois, Urbana-Champaign, September 1967.

(Received August 10, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 5, 439-444

VERBAL FACILITATION OF PAIRED-ASSOCIATE LEARNING: A LIMITED GENERALIZATION?[1]

JOEL R. LEVIN,[2] JAMES M. HORVITZ, AND SANDRA A. KAPLAN
University of Wisconsin

Two experiments were conducted in order to compare the sentence facilitation of paired associates under varying contexts. In Experiment I, 64 fifth graders were shown a 12-item list of printed noun pairs in one of four conditions, differing with respect to the kind of connective read aloud by the experimenter in relating the nouns (verb versus conjunction phrases), as well as the way in which the materials were visually presented to the subject (nouns only versus complete sentences or phrases). In Experiment II, an additional 16 fifth graders received all four utterance-presentation manipulations in a mixed list design. For both experiments, sentence facilitation of learning was obtained only when the sentences were heard, but not seen.

One of the most durable of psychological phenomena in recent years is the facilitative effect that syntactical mediation or verbal elaboration (Rohwer, 1967) has on learning. By now it is axiomatic that subjects who are provided with mediational strategies for organizing paired-associate, serial, or free-recall items will exhibit superior recall relative to those who are not. In fact, about the only hypotheses concerning verbal elaboration that are currently tested with regularity are directed toward *explaining* the effect, either in terms of the linguistic competencies of the populations under investigation (generally along a developmental or socioeconomic dimension) or in terms of the semantic and syntactic characteristics of the elaborative words or word strings themselves (e.g., Bobrow & Bower, 1969; Davidson & Dollinger, 1969; Ehri & Rohwer, 1969; Rohwer & Levin, 1968; Suzuki & Rohwer, 1968).

The remarkable consistency with which the verbal elaboration phenomenon has been demonstrated makes it all the more noteworthy that four recently completed experiments have failed to replicate the basic finding (Davidson, Schwenn, & Adams, 1970; Levin, 1970; Levin & Horvitz, 1971; Yuille & Pritchard, 1969). Although one might be tempted to suggest that the behavioral laws which pertain to Midwestern and Canadian elementary school children (the populations drawn from in these studies) are different than those which apply to Californians within the same age range (see Rohwer, 1967), an hypothesis more conducive to empirical validation is available upon a closer inspection of the four experiments.

In each of these studies it was found that noun pairs presented to subjects in the context of printed sentences (i.e., with elaboration) were no more easily learned than the same nouns presented by themselves or in the context of printed conjunctive phrases (without elaboration). In fact, in the Levin and Horvitz (1971) experiment, subjects who were shown noun pairs with printed elaboration actually performed at a significantly *lower* level than those who saw the pairs in conjunctive phrases.

Since it has previously been observed that the magnitude of elaboration effects from one experiment to the next is related to such variables as the kind of mediational strategy incorporated (experimenter-provided versus subject-generated), the method of testing adopted (recall versus recognition), the type of learning materials employed (verbal ver-

---

[1] This research was conducted while the first author was affiliated with Wisconsin Research and Development Center for Cognitive Learning with funds from the United States Office of Education, Department of Health, Education and Welfare, under the provisions of the Cooperative Research Program (Center No. C-03, Contract OE 5-10-154). The authors greatly appreciate the cooperation of the Stoughton, Wisconsin School District, and, in particular, the assistance of its Director of Instruction, Richard Carleton.

[2] Requests for reprints should be sent to Joel R. Levin, Research & Development Center for Cognitive Learning, 1404 Regent Street, University of Wisconsin, Madison, Wisconsin 53706.

sus pictorial) and so forth, it is possible that some variable or combination of variables shared by the studies just described is contributing to the elaboration washout.

In each of the four experiments, both the to-be-learned nouns and their companion elaboration were presented to subjects in printed form while simultaneously read aloud by the experimenter. Interestingly enough, while facilitation attributable to elaboration has been detected when paired associates are presented in this fashion (Rohwer, 1967), the effect is not nearly as dramatic as when paired *pictures* of objects are linked via sentences (e.g., Cooper, 1968; Jensen & Rohwer, 1963, 1965).

Thus, two obvious differences between heretofore maximally and minimally facilitative paradigms are identifiable: (*a*) the learning materials are printed words in the latter, and pictures in the former; and (*b*) the complete verbal context (i.e., the sentence) is both read and heard by the subject in the latter, but only heard by him in the former. With respect to the first of these differences, the comparative sentential facilitation for printed and pictorial stimuli has been reported elsewhere (Rohwer, Lynch, Levin, & Suzuki, 1967). In that experiment, sentences were found to facilitate the learning of both printed and pictorial paired associates, and to approximately the same degree.

Since no direct comparisons have been made regarding the form in which verbal elaboration is provided—read and heard by the subject versus just heard by him—two experiments were conducted to investigate possibly different sentence facilitation effects under the two conditions.

EXPERIMENT I

Method

Design and materials. Twelve pairs of familiar nouns were embedded either in sentences of the form article-noun-verb-article-noun, or in conjunctive phrases where the conjunction "and" or "or" replaced the verb of the sentence presentation. Conjunctive phrases, rather than the noun pairs by themselves, were employed as base-line materials in order to control for reading/study time differences in the two conditions. Furthermore, it has been demonstrated that paired associates embedded in conjunctive phrases produce recall equivalent to that obtained with traditional noun-noun learning groups (Rohwer, 1966; Rohwer, Shuell, & Levin, 1967).

Each pair of nouns was typed in capital letters on a separate 4 × 6 inch index card, with the second noun in each pair underlined. For half of the subjects, just the nouns themselves were typed on the cards while the accompanying sentence or phrase was read aloud by the experimenter (*pairs* only). For the other subjects, the complete sentence or phrase which incorporated the noun pairs was also typed on the index cards while read aloud by the experimenter (*complete* context). The first word of each string began with a capital letter, while the letters of words other than the two noun associates were all printed in lower case.

Experiment I, therefore, was a 2 × 2 design, with the two factors consisting of context (pairs versus complete) and connective (conjunction versus verb). Examples of the type of presentation supplied in each condition may be found in Table 1.

Procedure. All of the subjects were tested individually by two experimenters, one of whom served as a recorder of the subject's responses. Each subject had been randomly assigned to one of the four experimental conditions, and upon arriving at the testing room was given instructions appropriate for his condition. After informing the subject of his task, the experimenter presented three sample items followed by a clarification of the procedure based on the subject's indicated understanding of the directions.

Each of 12 noun pairs was then shown to the

TABLE 1

EXAMPLES OF THE FOUR EXPERIMENTAL CONDITIONS

Context	Connective	Subject	
		Sees	Hears
Pairs	Conjunction	GIRL *BALL*	The girl and the ball
	Verb	GIRL *BALL*	The girl grabs the ball
Complete	Conjunction	The GIRL and the *BALL*	The girl and the ball
	Verb	The GIRL grabs the *BALL*	The girl grabs the ball

subject at a 5-second rate, the experimenter exposing an index card while being paced by the appropriate verbalization on a tape recorder. Following the last study item was a 10-second interval, during which the experimenter prepared the subject for the test trial. Each first noun then appeared (on separate index cards) at a 5-second rate while named on the tape, with the subject's task being to vocally supply the missing underlined (second) noun. Another 10-second pause followed the last test item, and then the study-test cycle was repeated. Thus, a total of two study and two test trials was provided for each subject. Prior to conducting the actual experiment, three pilot subjects were run to help control for experimenter warm-up effects.

Subjects. Sixty-four subjects from the population of fifth graders at an elementary school serving a semirural community in the Midwest were randomly assigned (in equal numbers) to the four cells of the design, yielding a total of 16 subjects per experimental condition.

Results

Learning was measured by the number of response nouns correctly recalled by the subject on each of the two test trials. As there were only two dependent measures (number correct on each trial) for every subject, identical results would be obtained in conducting either "mixed-model univariate" (Winer, 1962) or "multivariate repeated measures" (Morrison, 1967) analyses of variance. Since the hypothesis of primary interest focused on the detection of sentential facilitation under each context type (pairs and complete), the connective comparisons were treated as being nested within contexts. All hypotheses were tested with the probability of a Type I error set equal to .05.

The mean number of items recalled over the two test trials by experimental condition, is presented in Table 2. The three between-group contrasts produced the following results: (*a*) No significant effect of context was found ($F < 1$); that is, pairs and complete context items were learned equally well. (*b*) A statistically significant verb-conjunction difference obtained when only the pairs were shown to the subject, that is, Connective within Pairs ($F = 4.80$, $df = 1/60$, $p < .05$); but (*c*) not when the complete context was provided, i.e., Connective within Complete ($F < 1$).

Although there was substantial improvement from Trial 1 to Trial 2 ($F = 301.06$, $df = 1/60$, $p < .0001$), no differential increment by experimental conditions was observed (i.e., the three Trials × Conditions interactions all produced F ratios with associated p values greater than .20).

TABLE 2

EXPERIMENT I: MEAN NUMBER OF CORRECT RESPONSES OVER TWO TRIALS BY EXPERIMENTAL CONDITION

Connective	Context		
	Pairs	Complete	Across contexts
Conjunction	11.94	13.31	12.62
Verb	14.94	13.06	14.00
Across connectives	13.44	13.19	

Note.—16 subjects per cell, maximum obtainable score is 24.

EXPERIMENT II

Method

In order to make a "generalizability check" on the results obtained in Experiment I, a second study was conducted employing the same experimental conditions, but this time manipulating them within subjects rather than between groups. By doing this, homogeneously and heterogeneously administered treatments may be compared (Levin, 1970) thereby providing valuable information with regard to the psychological reality of the initial findings.

Design and materials. Four 12-item mixed lists were constructed, each containing three items from the four experimental conditions of the previous experiment. That is, three items apiece from the "conjunction-pair," the "verb-pair," the "conjunction-complete," and the "verb-complete" conditions were presented to the subjects. Type of presentation was randomly interspersed throughout each of the four lists, with each list consisting of a particular item/presentation pairing as determined by a Latin square.

Procedure. The same procedure described previously was employed. The instructions and examples indicated to the subjects that the type of presentation would vary from item to item.

Subjects. A total of 16 additional subjects from the same population as in Experiment I was selected, and randomly assigned to the four lists, four subjects per list.

Results

For each subject, eight scores were recorded: the number of items correctly recalled under each of the four presentation

TABLE 3

EXPERIMENT II: MEAN NUMBER OF CORRECT RESPONSES OVER TWO TRIALS BY PRESENTATION METHOD

Connective	Context		
	Pairs	Complete	Across contexts
Conjunction	2.69	3.00	2.84
Verb	3.62	3.44	3.53
Across connectives	3.15	3.22	

Note.—16 subjects, maximum obtainable score is 6.

methods on the two test trials. A multivariate repeated measures analysis of variance (Morrison, 1967) was performed on the data, in which the same sources of variance as in Experiment I were examined. The mean number of correct responses over two trials, by presentation method, may be found in Table 3.

The results of the multivariate repeated measures analysis, that is, between presentation methods within subjects, were identical to the results of Experiment I. That is, sentence facilitation was detected when the elaboration was heard but not seen (the F ratio for Connective within Pairs was 5.15, $df = 1/15$, $p < .05$); but not when it was both read and heard by the subject (the F ratio for Connective within Complete was 1.21, $df = 1/15$, $p > .20$). Once again, the main effect of Context (pairs versus complete) was not significant ($F < 1$).

Finally, in agreement with the previous analysis, Trials was significant ($F = 240.25$, $df = 1/15$, $p < .0001$), but none of the interactions of Trials with the three presentation contrasts was (all Fs < 1).

DISCUSSION

The results of the present experiments indicate that the facilitation derived from verbal elaboration in children's paired-associate learning may be more limited (by task-specific variables) than is currently thought. In particular, a negligible effect of elaboration was observed here, as well as in some recent studies, when the auxiliary verbal material was printed along with the noun–noun associates.

At least three, perhaps nonmutually exclusive, explanations come to mind in accounting for the finding that a lesser degree of facilitation is produced when the verbal context is presented in printed, as opposed to aural, form. The first two of these explanations involve the notion of interference, one on a *physical* and the other on a *psychological* level.

The simplest hypothesis is that when the elaboration is printed, the verbs contained therein—which are seen by the subject—function as potential response options during recall. Stated simply: The more material read, the more material from which to select responses. If this hypothesis is correct, a greater number of verb-as-response errors would be expected to occur when the entire context was printed than when it was just heard. Inspection of the present data find this to be the case. (In the Davidson et al. [1970] study [with printed elaboration], it was also noted that previously seen verbs emerged as response errors.)

Apart from the greater number of physical opportunities for errors created by the printed verbs, it might also be argued that more associative or meaning responses are evoked when subjects read the elaboration. During recall, the subject may be able to retrieve the study verb, but then proceeds to pair it with a logical, though inappropriate, response. In this regard, there were 10% fewer "omission" errors and a greater proportion of "extralist" errors—among the overt errors—(34% versus 20%) in the printed, as compared with the heard, elaboration condition.

A third explanation suggests that a change in the utility of the mediator may be the consequence when the verbal elaboration is printed along with the nouns. Supplying the elaboration may convert a potentially effective learning strategy, for example, the use of meaningful sentences, into one where the subject is focusing his attention on the sequence of words without attending to their meaning. When subjects are required simply to listen to the elaboration, however, the comprehension or semantic element so vital to the sentence facilitation phenomenon (Bobrow & Bower, 1969; Rohwer, 1966; Rohwer & Levin, 1968) may be preserved to a greater degree.

It is important to note that the current series of nonreplications of the elaboration effect with printed materials has been localized among children in the primary grades. Others (e.g., Bower, 1968; Suzuki & Rohwer, 1969) have been able to demonstrate facilitation, even when the elaboration is printed, when college students serve as subjects. The developmental question, with regard to the kinds of presentations which are likely to be facilitative, is intriguing and deserves attention.

At the same time, after reviewing the rapidly accumulating research into verbal mnemonic systems, one might well propose a continuum of techniques ranging from extremely to mildly facilitative. On one end of the continuum would probably be found the subject-initiated strategies which include the learner's generation of unique verbal links with which to relate newly acquired information. Among the less potent facilitators would be already elaborated materials, where the learner is provided with appropriate mediators and therefore needs only *use*, rather than *generate*, them. Note that this kind of technique which appears to involve a less "active" information-processing on the learner's part, is also presumed to be less facilitative in terms of recall (see Bobrow & Bower, 1969, or Bower & Winzenz, 1970, for further discussion on this topic).

A parallel situation may be constructed for elaborative techniques in the visual domain, with certain strategies (e.g., subject-generated visual imagery) producing superior recall to others (e.g., learning from pictorial materials). It is interesting, however, that the limited generalizations which apply to elaboration in the verbal mode may have their counterpart in the visual as well.

For example, despite recent data which attest to the enormous learning gains produced by visual imagery (e.g., Bugelski, Kidd, & Segmen, 1968; Paivio, 1969; Persensky & Senter, 1969; Taylor & Black, 1969), Levin and Kaplan (in press) have found that this phenomenon, too, is not without restriction. Sixth graders who were given instructions and examples regarding the use of imagery prior to learning did markedly better than uninstructed subjects but—and here is the limitation—only when the learning materials consisted of pictured objects. When their noun referents were presented instead (as printed words), imagery instructions did not help the children. Thus, the encoding process as related to verbal and imaginal factors is critical, and needs to be systematically investigated along an age dimension.

Only after such experimentation has been completed will we be able to proffer more definitive statements regarding the efficacy of elaborative techniques. Of greater exigency, however, is the translation of such laboratory research findings into tangible—and hopefully, teachable—tactics which will have implications for school children, including those in California, Canada, and Wisconsin.

REFERENCES

BOBROW, S. A., & BOWER, G. H. Comprehension and recall of sentences. *Journal of Experimental Psychology*, 1969, **80**, 455–461.

BOWER, G. H. Organizational factors in memory. Research Seminar, Institute of Human Learning, University of California, Berkeley, 1968.

BOWER, G. H., & WINZENZ, D. Comparison of associative learning strategies. *Psychonomic Science*, 1970, **20**, 119–120.

BUGELSKI, B. R., KIDD, E., & SEGMAN, J. Image as a mediator in one-trial paired-associate learning. *Journal of Experimental Psychology*, 1968, **76**, 69–73.

COOPER, C. S. The relationships among reading ability, grade level, and syntactical mediation in paired-associate learning. Unpublished master's thesis, University of Wisconsin, 1968.

DAVIDSON, R. E., & DOLLINGER, L. The syntactic facilitation of paired-associate learning. *Journal of Educational Psychology*, 1969, **60**, 434–438.

DAVIDSON, R. E., SCHWENN, E. A., & ADAMS, J. F. Semantic effects in transfer. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 212–217.

EHRI, L. C., & ROHWER, W. D., JR. Verb facilitation of paired-associate learning as a function of syntactic and semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 773–781.

JENSEN, A. R., & ROHWER, W. D., JR. The effect of verbal mediation on the learning and retention of paired-associates by retarded adults. *American Journal of Mental Deficiency*, 1963, **68**, 80–84.

JENSEN, A. R., & ROHWER, W. D., JR. Syntactical mediation of serial and paired-associate learning as a function of age. *Child Development*, 1965, **36**, 601–608.

LEVIN, J. R. Note on mixed lists in the analysis of repeated measurements. Technical Memo No. S-70-2, Wisconsin Research and Development

Center for Cognitive Learning, Madison, July 1970.

LEVIN, J. R. Factors related to the sentence facilitation of paired-associate learning: Some characteristics of verbs. *Journal of Educational Psychology*, 1970, 61, 431–439.

LEVIN, J. R., & HORVITZ, J. M. The meaning of paired associates. *Journal of Educational Psychology*, 1971, 62, 209–214.

LEVIN, J. R., & KAPLAN, S. A. Imaginal facilitation of paired-associate learning: A limited generalization? *Journal of Educational Psychology*, in press.

MORRISON, D. F. *Multivariate statistical methods*. New York: McGraw-Hill, 1967.

PAIVIO, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, 76, 241–263.

PERSENSKY, J. J., & SENTER, R. J. An investigation of 'bizarre' imagery as a mnemonic device. Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago, May 1969.

ROHWER, W. D., JR. Constraint, syntax, and meaning in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 541–547.

ROHWER, W. D., JR. Social class differences in the role of linguistic structures in paired-associate learning: Elaboration and learning proficiency. (Final report on United States Office of Education Basic Research Project No. 5-0605, Contract No. OE 6-10-273) Washington, D.C.: United States Government Printing Office, 1967.

ROHWER, W. D., JR., & LEVIN, J. R. Action, meaning, and stimulus selection in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1968, 7, 137–141.

ROHWER, W. D., JR., LYNCH, S., LEVIN, J. R., & SUZUKI, N. Pictorial and verbal factors in the efficient learning of paired associates. *Journal of Educational Psychology*, 1967, 58, 278–284.

ROHWER, W. D., JR., SHUELL, T. J., & LEVIN, J. R. Context effects in the initial storage and retrieval of noun pairs. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 796–801.

SUZUKI, N., & ROHWER, W. D., JR. Verbal facilitation of paired-associate learning: Type of grammatical unit vs. connective form class. *Journal of Verbal Learning and Verbal Behavior*, 1968, 7, 584–588.

SUZUKI, N., & ROHWER, W. D., JR. Deep structure in the noun-pair learning of children and adults. *Child Development*, 1969, 40, 911–919.

TAYLOR, A. M., & BLACK, H. B. Variables effecting imagery instruction in children. Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago, May 1969.

WINER, B. J., *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

YUILLE, J. C., & PRITCHARD, S. Noun concreteness and verbal facilitation as factors in imaginal mediation and PA learning in children. *Journal of Experimental Child Psychology*, 1969, 7, 459–466.

(Received August 18, 1970)

ERRATUM

In the article by Thomas J. Quirk, Margaret T. Steen, and Dewey Lipe, in the June 1971 issue, a misprint occurred in Table 3, page 197. The correct value for *z* in the last column on the right for Category 12 should be .25 and not 6.25.

from HARPER & ROW

JUST PUBLISHED & FORTHCOMING . . .

Bigge

LEARNING THEORIES FOR TEACHERS, SECOND EDITION

MORRIS L. BIGGE. Treating prescientific and scientific theories of learning, this book helps the reader to critically analyze existing theories and then guides him in formulating his own views on learning and in deciding how best to implement these views. The Second Edition—extensively revised, expanded, and more readable—relates learning theories to teaching procedures throughout. Readers are led to making a choice among competitive psychological points of view—strengths of each theory are pointed out, but it is left to the reader to discern the weaknesses. Included is a clear discussion of such familiar topics as transfer of learning, motivation, retention, review, evaluation, practice vs. repetition, and part vs. whole learning. October, 1971. Paper. 416 pp. (tentative); $4.95.

Dreikurs • Grunwald • Pepper

MAINTAINING SANITY IN THE CLASSROOM

ILLUSTRATED TEACHING TECHNIQUES

RUDOLF DREIKURS, BERNICE BRONIA GRUNWALD, and FLOY C. PEPPER. Based on Adlerian psychology and heavily case-oriented, this text applies techniques of *motivation modification* to classroom problems. Pairing effective teaching with effective leadership, the authors systematically utilize their own case material to illustrate teaching problems and how to deal with them. Covering both intergroup and interpersonal conflicts, the text stresses group control of behavior through group participation and cooperation and provides a wealth of solutions to classroom problems and suggestions for successful teaching. October, 1971. Paper. Tentative: 350 pp.; $4.95.

Greenwood • Good • Siegel

PROBLEM SITUATIONS IN TEACHING

GORDON E. GREENWOOD, THOMAS L. GOOD, and BETTY L. SIEGEL. Utilizing a decision-making view of teaching, this book includes 20 representative case studies, developed from data on over 330 actual problem cases. Prefaced by a detailed introduction that suggests a strategy for attacking teaching problems, the cases, each reported behaviorally and left unresolved, allow for original student analysis and solutions. Questions following each case serve to provoke and guide discussion. Written to function in place of, or to supplement, films, video tapes, and other media, the book is adaptable to a wide range of courses and to a variety of teaching methods. Instructor's Manual. Paper. September, 1971. 224 pp. (tentative); $3.95.

Vargas

WRITING WORTHWHILE BEHAVIORAL OBJECTIVES

JULIE S. VARGAS. Teaching how to recognize and classify behavioral objectives and how to write quality objectives *at every cognitive level*, this self-instructional text models the principles it teaches—statement of objectives, pretests, graded exercises with answers, and posttests. Unique to this book is its behavioral analysis of concept formation and creativity and its description of how to incorporate them into a course. January, 1971. Paper. Tentative: 160 pp.; $2.95.

Turn the page for descriptions of other Harper & Row texts of interest

1817

from HARPER & ROW

JUST PUBLISHED . . .

Banaka

TRAINING IN DEPTH INTERVIEWING

WILLIAM H. BANAKA. Here is a succinct training manual emphasizing an interpersonal communications theory (adapted from Bales and Schutz) and rigorous self-evaluation. It is the first manual to provide highly detailed criteria and guidelines for intensive evaluation of taped (audio or video) interviews and to incorporate sensitivity techniques into its training methods. All facets of the information-gathering interview are systematically covered, with theory and evaluation procedures interwoven throughout the text. The manual is readily adaptable to an individual instructor's specific type of interview and course objectives. October, 1971. Paper. Tentative: 128 pp.; $1.95.

OTHER VALUABLE SELECTIONS . . .

Cronbach

ESSENTIALS OF PSYCHOLOGICAL TESTING, THIRD EDITION

LEE J. CRONBACH. 1970. 752 pp.; $10.95.

Downie • Heath

BASIC STATISTICAL METHODS, THIRD EDITION

N. M. DOWNIE and ROBERT W. HEATH. 1970. 356 pp.; $10.50/*Study Guidebook:* $3.50.

Maslow

MOTIVATION AND PERSONALITY, SECOND EDITION

ABRAHAM H. MASLOW. 1970. Paper. 369 pp.; $6.50.

Mink

THE BEHAVIOR CHANGE PROCESS

OSCAR G. MINK. 1970. Paper. 208 pp.; $4.25.

Mussen • Conger • Kagan

CHILD DEVELOPMENT AND PERSONALITY, THIRD EDITION

PAUL HENRY MUSSEN, JOHN JANEWAY CONGER, and JEROME KAGAN. 1969. 795 pp.; $11.50/*STUDY GUIDE* by Fay-Tyler Norton: $2.95.

READINGS IN CHILD DEVELOPMENT AND PERSONALITY,
SECOND EDITION

MUSSEN, CONGER, and KAGAN. 1970. Paper. 595 pp.; $6.50.

Oetting • Thornton

EXERCISES IN PSYCHOLOGICAL TESTING

EUGENE R. OETTING and GEORGE C. THORNTON III. 1968. Paper. 229 pp.; $4.95.

Staats

CHILD LEARNING, INTELLIGENCE, AND PERSONALITY
PRINCIPLES OF A BEHAVIORAL INTERACTION APPROACH

ARTHUR W. STAATS. 1971. 358 pp.; $8.95.

1817

For more information on these and other Harper & Row texts in the field of psychology write Department 275 (E/1); for more information on our education texts write Department 275 (E/2).

from HARPER & ROW

A RECENTLY RELEASED EDITION OF A CLASSIC IN THE FIELD . . .

Klausmeier • Ripple

LEARNING AND HUMAN ABILITIES

EDUCATIONAL PSYCHOLOGY, THIRD EDITION

HERBERT J. KLAUSMEIER and RICHARD E. RIPPLE. This leading text presents a systematic unification of the field from the perspective of its primary concerns: the nature, conditions, outcomes, and evaluation of classroom learning. With a realistic awareness of how educational practices can best be applied today, the Third Edition takes into account the negative effects on education of war, drugs, violence, racism, and poverty. Completely rewritten for clarity and interest, the text incorporates the results of the latest empirical research studies. Among the organizational changes: individual differences, instructional materials, and motivation are covered earlier in the discussion; factual information and concepts, formerly one chapter, are now treated in two separate expanded chapters. Internal study aids include an advance organizer (outline of key concepts) for each chapter, questions for review and discussion at the end of each main section of each chapter, and "flagging" (by brackets) of important within-text passages. Profusely illustrated. *File of Test Items* by Ruth S. Nickse and Richard E. Ripple. 1971. 810 pp.; $12.00/*STUDENT GUIDE* by Ruth S. Nickse and Richard E. Ripple (contains, for each chapter in the text, an outline of the chapter, a multiple-choice self-test, discussion questions, and projects; also, answers to all self-tests, supplementary reading list of general books, and annotated list of professional journals): $4.00.

TO SUPPLEMENT ANY INTRODUCTORY EDUCATIONAL PSYCHOLOGY TEXT . . .

Ripple

READINGS IN LEARNING AND HUMAN ABILITIES

EDUCATIONAL PSYCHOLOGY, SECOND EDITION

RICHARD E. RIPPLE. This exceptional collection of 41 unabridged readings—most of them dealing with empirical research—introduces the student to original literature in the field. Three-quarters of the material postdates publication of the First Edition. Sectional introductions; interpretive headnote for each reading. 1971. Paper. 484 pp.; $6.50.

ALSO OF INTEREST . . .

Gordon

HUMAN DEVELOPMENT

FROM BIRTH THROUGH ADOLESCENCE, SECOND EDITION

IRA J. GORDON. 1969. 408 pp.; $10.00.

1817

Harper & Row, Publishers, Inc.
49 E. 33d Street, N.Y. 10016

Journal of the Experimental Analysis of Behavior

The *Journal of the Experimental Analysis of Behavior* is primarily for the original publication of experiments relevant to the behavior of individual organisms. Review articles and theoretical papers will also be considered for publication. Invited book reviews appear on occasion.

EDITOR

STANLEY S. PLISKOFF, *University of Maine*

EXECUTIVE EDITOR

VICTOR G. LATIES, *Univ. of Rochester*

ASSOCIATE EDITORS

LARRY D. BYRD, *Harvard Medical School*
LEWIS R. GOLLUB, *Univ. of Maryland*
HERBERT M. JENKINS, *McMaster Univ.*
HOWARD RACHLIN, *SUNY, Stony Brook*
MICHAEL D. ZEILER, *Emory University*

REVIEW EDITOR

A. CHARLES CATANIA, *New York Univ.*

BOARD OF EDITORS

W. BAUM, *Harvard University*
D. S. BLOUGH, *Brown University*
J. J. BOREN, *American University*
F. C. CLARK, *Univ. of North Carolina*
M. R. D'AMATO, *Rutgers University*
J. A. DINSMOOR, *Indiana University*
E. FANTINO, *Univ. of Cal., La Jolla*
G. W. FARTHING, *University of Maine*
R. M. GILBERT, *Alcoholism & Drug Addiction Res. Foundation*
D. F. HAKE, *Anna State Hospital*
R. F. HEFFERLINE, *Columbia University*
R. M. HERRICK, *Air Dev. Ctr., Johnsville*
P. N. HINELINE, *Temple University*
W. HODOS, *Walter Reed*
W. C. HOLZ, *Smith Kline & French*
D. JOHNSON, *Virginia Polytec. Inst.*
R. T. KELLEHER, *Harvard Medical Sch.*
D. KELLY, *NYS Psychiatric Institute*
J. W. MCKEARNEY, *Worcester Foundation*
M. J. MARR, *Georgia Inst. of Tech.*
J. R. MILLENSON, *Univ. of Oxford*
D. E. MINTZ, *CUNY, City College*
W. H. MORSE, *Harvard Medical School*
A. NEURINGER, *Reed College*
B. RAY, *Mass. General Hospital*
S. H. REVUSKY, *N. Illinois University*
G. S. REYNOLDS, *Univ. of Cal., La Jolla*
B. SCHNEIDER, *Columbia University*
J. SHERMAN, *University of Kansas*
C. SHIMP, *University of Utah*
J. E. R. STADDON, *Duke University*
D. A. STUBBS, *University of Maine*
H. S. TERRACE, *Columbia University*
T. VERHAVE, *CUNY, Queens College*
B. WEISS, *Univ. of Rochester*
D. R. WILLIAMS, *Univ. of Pennsylvania*
M. YARCZOWER, *Bryn Mawr College*
J. ZIMMERMAN, *Indiana Univ. Med. School*

During 1971 the following articles of interest to sensory psychologists have already appeared in JEAB:

Beecher, M. D. and Harrison, J. M. Rapid acquisition of an auditory localization discrimination by rats.

Blough, Patricia M. The visual acuity of the pigeon for distant targets.

Blue, S., Sherman, J. G. and Pierrel, R. Differential responding as a function of auditory stimulus intensity without differential reinforcement.

Ernst, A. J., Engberg, L. and Thomas, D. R. On the form of stimulus generalization curves for visual intensity.

Harrison, J. M., Downey, P., Segal, M. and Howe, M. Control of responding by location of auditory stimuli: Rapid acquisition in monkey and rat.

Moody, David B., Stebbins, William C., and Iglauer, Carol. Auditory generalization gradients for response latency in the monkey.

Siegel, R. K. Apparent movement and real movement detection in the pigeon: Stimulus generalization.

Williams, Ben A. Color alternation learning in the pigeon under fixed-ratio schedules of reinforcement.

Wilton, R. N. and Clements, R. O. Observing responses and informative stimuli.

Winton, A. S. W. and Beale, I. L. Peak shift in concurrent schedules.

Wright, Anthony A. and Cumming, William W. Color-naming functions for the pigeon.

Address subscription orders to Kay Dinsmoor, JEAB

Department of Psychology
Indiana University
Bloomington, Indiana 47401

Students $4 (payment with order)—Individuals (personal use) $8
Institutions $16

ECEA Pinpoints More Information on What's Being Published in Special Education Than Any Other Reference

Maybe you're researching learning disabilities. Or new methods for teaching the deaf. Or driver education for the mentally handicapped. Or the emotionally disturbed.

You get this coverage and more with *Exceptional Child Education Abstracts* . . . the single most comprehensive guide to current literature on education for exceptional children, both handicapped and gifted.

Over 3,000 abstracts are published annually, spanning a wide range of subjects and disciplines . . . an invaluable research tool for librarians, teachers, curriculum supervisors, administrators, students.

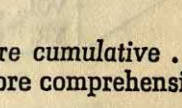
ECEA speeds research time because *computer generated indices are cumulative* . . . by author, subject and title. The researcher can quickly identify more comprehensive information than would be possible by any other method.

Many documents abstracted are available in microfiche or hard copy at nominal cost.

Get to know the now way to know . . . ECEA, a product of the CEC Information Center (CEC-ERIC Clearinghouse). Published quarterly by The Council for Exceptional Children.

EXCEPTIONAL CHILD EDUCATION ABSTRACTS

TO: The Council for Exceptional Children
Department 130
1411 South Jefferson Davis Highway
Arlington, Virginia 22202



For APA Members . . . special individual rate, $35 per year
(Regular Institutional Rate is $50)

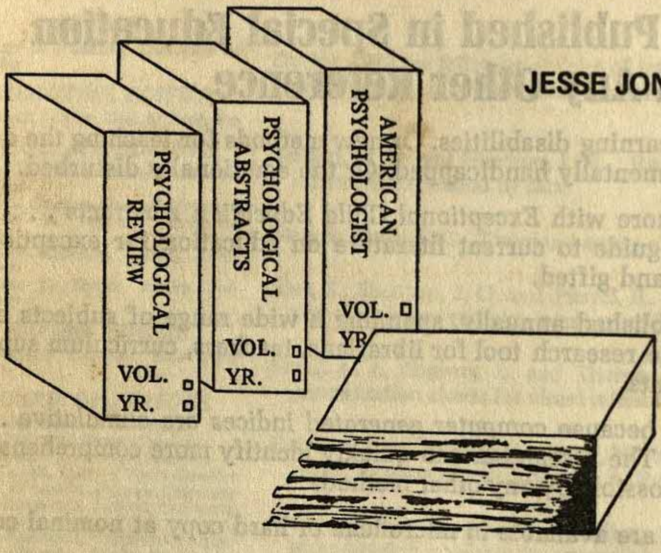
Please enter my order for subscription(s) to ECEA, Vol. III (4 issues)

☐ Check Enclosed. My P.O. No. is ____ ☐ Please bill me.
☐ Send descriptive brochure.

Institution____________________

Name____________________

Address____________________
City State Zip



JESSE JONES VOLUME FILES

Send orders to:
JESSE JONES CORP.
P.O. Box 5120
Philadelphia, Pa. 19141

Only $3.50 ea.
3 for $10.00
6 for $19.00 P.P.D.

Now, the journals of the American Psychological Association can become an attractive permanent part of your professional library. These famous Jesse Jones volume files, especially designed to keep your copies orderly, readily accessible for future references — guard against soiling, tearing, wear or misplacement of copies.

This durable file will support 150 lbs. Looks and feels like leather and is washable. The 23-carat gold lettering makes it a fit companion for the most costly binding.

JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 62 December 1971 Number 6

CONTENTS

Influence of Fading Procedures on the Von Restorff Effect 445
STEVEN M. JUNG

Effect of Feedback Expectancy on Choice of Teaching Styles 451
EDMUND T. EMMER, THOMAS L. GOOD, AND THOMAS D. OAKLAND

Reexamination of Two Learning Style Studies in the Light of the Cognitive Process Theory of Curiosity 456
DAVID G. BESWICK AND G. KASTEN TALLMADGE

Interaction of Anxiety with Performance on Two Levels of Task Difficulty 463
ROBERT D. TENNYSON AND F. ROSS WOOLLEY

Divergent Thinking in Blind Children 468
WILLIAM J. TISDALL, A. EDWARD BLACKHURST, AND CLAUDE H. MARKS

Response Latency: Errors and Subsequent Change of Response on Retest 474
PHYLLIS W. BERMAN AND MARY ANN VILLWOCK

Auditory-Visual Integration, Sex, and Reading Achievement 482
DAVID H. REILLY

Student Imitation of a Rewarding Teacher's Verbal Style as a Function of Sex and Grade Level 487
PHILIP FRIEDMAN AND NORMAN D. BOWERS

Perceptual Speed in Poor Readers 492
CARL SPRING

Visual Search and Reading 501
ROBERT C. CALFEE AND PENNY JAMESON

Comprehension Training and Identification for Poor Readers and Good Readers 506
ROLLIN STEINER, MORTON WIENER, AND WARD CROMER

Structural Coherence in Pictorial and Verbal Displays 514
RONALD W. SPANGENBERG

Effect of Sentence-Combining Exercises upon Several Restricted Written Composition Tasks 521
M. R. VITALE, F. J. KING, D. W. SHONTZ, AND G. M. HUNTLEY

Imagery and Sentence Learning 526
RICHARD C. ANDERSON AND JANET L. HIDDE

Concept Attainment of Culturally Advantaged and Disadvantaged Children Utilizing Artificial and Lifelike Stimulus Tasks 531
SAMUEL SECURRO, JR. AND RICHARD T. WALLS

This is the last issue of Volume 62
Volume Title Page and Volume Contents appear herein.

© 1971 by the American Psychological Association, Inc.

Journal of Educational Psychology
1971, Vol. 62, No. 6, 445–450

INFLUENCE OF FADING PROCEDURES ON THE VON RESTORFF EFFECT

STEVEN M. JUNG
American Institutes for Research, Palo Alto, California

This study investigated the influence of fading procedures on transfer of learning which resulted from single-item isolation by color. Sixty subjects were utilized in a multifactor experiment. The first factor was represented by two levels of pretraining, in which subjects read paragraphs which were relevant and irrelevant to an isolated stimulus. The second factor was represented by three levels of stimulus isolation during paired-associate learning. The first level was a control condition in which all stimuli were uncolored. In the second level, one stimulus was consistently projected in red. The third level was a fading condition in which the color was faded out over the course of repeated trials. Results demonstrated that significantly fewer trials were required to learn an association to the stimulus under isolated and fading conditions. This finding was reversed on a transfer measure. Control subjects gave significantly more correct responses to the stimulus on color-removed trials than did isolated condition subjects with fading condition subjects performing at an intermediate level.

Researchers in the field of reading have attempted to facilitate student learning by producing reading primers with words and syllables printed in distinctive colors (Jones, 1965; Kjeldergaard & Frankenstein, 1967; Otto, 1967, 1968). Samuels (1968) has called these attempts into question by demonstrating that although single-item isolation by color produces positive effects during the learning of highly similar materials, the effects are reversed in a transfer situation in which the artificial distinctiveness is removed.

This demonstration is analogous to the findings of Anderson and his collaborators (Anderson & Faust, 1967; Anderson, Faust, & Roderick, 1968) who found that overprompting in paired-associate learning or in programmed instruction resulted in definite performance decrements on posttraining examinations, although fewer errors were made during training.

Apparently there is a tendency for learners to focus on a stimulus dimension which, although serving reliably as a cue during learning, is removed or made nonfunctional at transfer time. This is especially the case when it is difficult to discriminate among the stimulus materials under ordinary circumstances.

There is some question about Samuels' (1968) contentions relating the von Restorff isolation effect to color cueing in reading. An analogy can probably not be drawn. Pillsbury and Raush (1943) have shown a significant decrement in learning as the number of isolated items in a list is increased. This suggests that as habituation occurs, or distinctiveness is compromised, there is less of an isolation effect. The implication is that continued use of colors in teaching a sight vocabulary would soon exhaust any possible manifestation of the von Restorff effect. Nevertheless, having demonstrated that transfer is hindered following single stimulus isolation, Samuels would presumably make the same prediction in the case of a more defensible application, such as in attention focusing through increased distinctiveness.

Anderson and his collaborators have made a similar contention relating to transfer failure in the context of programmed instruction. Again, however, an analogy between overprompting and the von Restorff effect may be made only cautiously.

The present study was designed to investigate the previous finding that use of attention-getting devices to enhance learning results in performance decrements on transfer tasks. A fading procedure was employed to test the prediction that gradual elimination of isolating cues during training would reverse and transfer decrement.

Taber and Glaser (1962) have demonstrated the effectiveness of cue fading during training in a case in which a transfer of discrimination was to be taught. It is of note that although Samuels (1968) and other authors (Anderson et al., 1968; Duell, 1968; Otto, 1968) have called the use of artificially distinctive cues into question because of their failure at transfer time, they have offered subjects limited opportunity to learn under conditions in which the irrelevant stimulus aspects are *gradually* faded out. The process of interspacing transfer trials among learning trials (Samuels, 1968) may be equivalent to a fading procedure, since it presumably allowed subjects to realize that color isolation was being removed. The present study was proposed as a test of this equivalence. Should fading of isolation in a von Restorff situation not prove detrimental to transfer, the educational use of such cues under proper circumstances could be more strongly defended.

In summary, the present study examined performance as measured by correct anticipations on a paired-associate learning task under various conditions of single stimulus isolation. Isolation was achieved by manipulation of prior training under various degrees of artificial color distinctiveness. The effect of artificial distinctiveness fading on transfer was then examined. It was hypothesized that fading would eliminate a transfer decrement produced by a nonfading condition.

METHOD

Subjects

Subjects were 60 undergraduate students who volunteered to participate as part of their involvement in a senior level educational psychology course. The nature of the experiment was explained during the first week of the course. Volunteers were asked to sign a participation sheet indicating the time at which they would appear. At the time each subject signed the participation sheet, he took a reminder slip and wrote the indicated time on it. The reminder slip also contained directions to help subjects find the experimental site.

Subjects ranged in age from 19 to 55 years. A sex breakdown of approximately 50-50 was achieved. No subjects reported prior familiarity with the particular nature of the experiment.

Two subjects were unable to master the paired-associate learning task. One subject was in the fading-of-isolation group and one was a control. These subjects were excused and replaced by two other volunteers.

Apparatus

The apparatus consisted of a Kodak Ectagraphic slide projector, three 80-slide Carousel trays, and a 5 × 7 inch slide-projection screen. Three sets of slides were prepared and inserted in the trays, one set and tray representing each of the three presentation conditions. In addition, two pretraining paragraphs were constructed, each representing one level of prior relevancy training.

Seven Greek stimulus words were constructed by selecting three Greek characters at random with the restrictions that three different characters were used in each word and no word started or ended with the same character. The seven response words were low-meaning English letter trigrams adapted from Samuels (1968) Experiment II. Stimuli and responses were then paired randomly. The completed paired-associate list was composed of the seven pairs indicated in Table 1.

Training materials. Slides were prepared using a photographic process. Lower case press-on lettering was used in all training materials. The lettering was transferred to a sheet of acetate film which had been marked off in the shape of 2 × 2 inch slides. The prepared acetate sheet was then placed over sheets of 8 × 10 inch film and exposed to a light source for 30 seconds under darkroom conditions. The negatives were developed, dried overnight, and cut into 2 × 2 inch slides which were mounted in standard mounts. The above process produced slides which projected white letters on a solid black background.

In order to achieve the red coloring essential to the two isolated-stimulus training conditions, carmen water color was used. The all-red slides were painted on the emulsion side with five coats of water color. The first-fade red stimulus received three coats, while the final-fade pink stimulus received one coat. This produced a progression of slides which projected first in bright red, then in light red, and finally in white with a barely visible pink tinge, all on the standard black background.

In order to preclude any serial learning effects, three random arrangements of each list were produced for each of the three treatment conditions.

TABLE 1

STIMULUS AND RESPONSE PAIRS UTILIZED IN ALL TREATMENT CONDITIONS

Pair number	Stimulus term	Response term
1	ηγβ	vuf
2	φθη	uaz
3	ψνλ	fvy
4	ωρτ	zyx
5	σλκ	xyz
6	λοφ	xaz
7	χδα	zof

These were followed by one random arrangement of only the stimulus terms for the transfer test. In all, 157 slides were constructed, 49 for each treatment. The completed sets were placed in three Kodak Carousel trays. The stimulus term of Pair 4 was colored under the two isolation training conditions, but it remained white in the control training condition and in all transfer tests.

Pretraining materials. Two paragraphs were prepared to be read by subjects prior to learning the paired associates. One paragraph, the relevant passage, was constructed so as to focus attention on the three Greek-character components of Stimulus Term 4. It contained short explanations of the Spearman rho, Kendall tau, and Kendall W correlation techniques. The exact characters which composed the fourth stimulus were included as headings for each explanation. A second paragraph, the irrelevant passage, was not related to the learning task. It consisted of a short exposition on computer programming languages.

Procedure

Upon arrival in the experimental setting, subjects were assigned at random to one of the two pretraining conditions and one of the three training conditions. Each subject thus came into one of the six cells of the basic 2 × 3 multifactor design. Subjects were seated facing the apparatus and were given a set of instructions regarding the experiment. They were then given either the relevant or the irrelevant pretraining passage and asked to simply read through it. During this time, the experimenter placed the Carousel tray containing the slides appropriate to the chosen training condition onto the projector.

Slides for the control training condition all projected white on black. Slides for the isolated-no-fading condition were the same except that the stimulus term of Pair 4 projected in dark red; the stimulus term of Pair 4 in the fading-of-isolation condition was dark red on the first presentation, light red on the second, and faintly pink on the third.

Immediately upon completion of the pretraining passage, subjects were allowed to preview the second arrangement of the paired-associate list in the slide tray. This was done to prevent any serial learning that may have occurred had the first arrangement been used for anticipation training twice in succession. Stimulus and response slides were each presented for a 5-second period. The tray was then returned to zero point and subjects were told to begin anticipating out loud any time they felt they knew the English letters which would follow each Greek stimulus. The three random list arrangements in the tray were presented automatically, constituting one learning trial block. A blank slide separated each arrangement. Stimuli only were then presented and subjects were instructed that the experimenter desired to see how many associations they had learned up to that point. During this transfer test, the experimenter took over manual control of the slide projector using a remote cord to ensure sufficient time for a response to each stimulus. Following the transfer test, the slide tray was manually returned to zero point again and a new learning trial block was begun. Three trial blocks and three transfer tests constituted the entire experiment.

A record of all correct anticipations was kept during the course of the nine training and three transfer trials. These were tabulated giving total correct anticipations per training block for each subject. In addition, correct responses to Stimulus 4, the isolated stimulus, were tabulated for each training block and for the total of three transfer tests. These data served as the dependent measures for the study.

RESULTS

The data generated by the present experiment were analyzed in three major parts. The first part is a multifactor analysis of variance with repeated measures (Winer, 1962) using number of correct anticipations to the isolated stimulus as the dependent variable. The second part is a two-way analysis of variance using total correct transfer responses as the dependent variable. The final analysis was performed in the same manner as the first, using total correct anticipations during training as the dependent variable.

The null hypothesis for the first analysis stated that there would be no significant effects for pretraining, fading procedures, or trial blocks on the ease with which an association to an isolated or control stimulus is acquired. Table 2 presents the cell means and standard deviations and Table 3 presents the marginal means of main treatment

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR CORRECT ANTICIPATIONS TO ISOLATED STIMULUS

Pretraining	Fading procedures	Trial blocks 1		2		3	
		M	*SD*	*M*	*SD*	*M*	*SD*
Irrelevant	Control (No isolation)	.10	.30	1.10	.83	1.70	1.00
	Isolated (No fading)	1.20	1.16	2.30	.90	2.50	.92
	Fading of isolation	1.40	.92	1.90	1.04	2.40	1.02
Relevant	Control (No isolation)	.90	.70	2.00	.63	2.20	1.25
	Isolated (No fading)	1.70	.90	2.00	1.09	2.70	.46
	Fading of isolation	1.50	.92	2.00	.89	2.70	.64

Note.—$N = 60$; 10 per cell.

TABLE 3

MARGINAL MEANS FOR CORRECT ANTICIPATIONS TO ISOLATED STIMULI

	M
Pretraining	
Irrelevant	1.62
Relevant	1.82
Fading procedures	
Control	1.28
Isolated	1.90
Fading	1.98
Trial blocks	
1	.97
2	1.90
3	2.33

effects for correct associations to the isolated stimulus.

Isolation procedures had a significant effect on number of correct responses to the isolated item ($F = 5.21$, $df = 2/108$, $p < .05$). In addition, learning trials had a highly significant effect ($F = 61.88$, $df = 2/108$, $p < .01$). From an examination of Table 3, it would appear that these results are due to easier learning of the isolated item over its uncolored counterpart and to an expected improvement in performance with repeated learning trials. This may be taken as a demonstration of the classical von Restorff effect. No evidence of a pretraining effect or significant Isolation × Trial interaction for correct isolated item anticipations was found.

The means, standard deviations, and marginal means for transfer performance are presented in Table 4.

It appears that although the group which received a color-isolated stimulus acquired a response to that stimulus more rapidly, transfer performance was adversely affected. Fading procedures produced only a slight improvement in this situation. Control subjects performed significantly better on the transfer task ($F = 3.17$, $df = 2/54$, $p < .05$). Again, no significant effects for pretraining procedures or interactions were found on the transfer measure.

TABLE 4

MEANS, STANDARD DEVIATIONS, AND MARGINAL MEANS FOR TOTAL TRANSFER RESPONSES

Pretraining	Isolation procedures						Marginal *M*
	Control		No fading		Fading		
	M	*SD*	*M*	*SD*	*M*	*SD*	
Irrelevant	1.60	.71	.70	.89	1.40	.78	1.20
Relevant	1.80	.62	1.10	.70	1.40	.69	1.43
Marginal *M*	1.70		.90		1.40		

Note.—$N = 60$; 10 per cell.

TABLE 5

MEANS AND STANDARD DEVIATIONS FOR TOTAL CORRECT ANTICIPATIONS

Pretraining	Fading procedures	Trial blocks					
		1		2		3	
		M	*SD*	*M*	*SD*	*M*	*SD*
Irrelevant	Control (No Isolation)	1.70	1.19	5.50	2.37	9.40	3.80
	Isolated (No fading)	4.00	2.86	7.50	3.38	12.40	4.32
	Fading of isolation	4.30	2.15	7.90	2.55	12.50	3.17
Relevant	Control (No isolation)	3.20	.98	6.50	2.87	10.00	3.82
	Isolated (No fading)	3.50	1.36	8.10	2.43	12.70	3.03
	Fading of isolation	4.30	2.53	6.50	4.13	11.30	5.40

Note.—$N = 60$; 10 per cell.

TABLE 6

MARGINAL MEANS FOR TOTAL CORRECT ANTICIPATIONS

	M
Pretraining	
Irrelevant	7.23
Relevant	7.35
Fading procedures	
Control	6.05
Isolated	8.03
Fading	7.80
Trial blocks	
1	3.50
2	7.00
3	11.38

Total correct anticipations were investigated in the third analysis. Table 5 presents the means and standard deviations for this variable; the marginal means of main treatment effects are presented in Table 6.

Only the effect of trials within subjects was significant ($F = 224.07$, $df = 2/108$, $p < .01$). The main effect for isolation procedures, which had been significant in the analysis of correct responses to the isolated item, was not significant in the analysis of total correct anticipations ($F = 2.97$, $df = 2/108$, $p > .05$).

DISCUSSION

It was originally hypothesized that fading techniques would serve to overcome decrements in isolated stimulus transfer performance noted by Samuels (1968). He found that removal of color isolation following the production of a classical von Restorff effect resulted in confusion upon presentation of the previously isolated stimulus. His findings are replicated by the present results. However, the predicted role of fading failed to materialize. On learning measures, the faded isolated stimulus served almost equally as well as the nonfaded stimulus, and both served significantly better than the noncolored control stimulus. On transfer measures, the faded stimulus evoked more correct responses than the nonfaded, but less than the control stimulus.

Apparently the fading-of-color technique was still insufficient to cause learners to seek out the actual cues of the Greek stimulus word. Subjective observations by the experimenter confirm this. When the pink version of the isolated stimulus was presented, many subjects who later failed the transfer test would exclaim, "Oh, that's the red one." Samuels' original objections to any serious educational application of the von Restorff phenomenon are supported.

The theoretical implications of this failure of fading may be pointed out here. If unexpected change in stimulation actually causes some type of physiological orienting reflex which results in increased learning, any perceivable change of lesser magnitude should also have this effect. Since the reaction would be in the central nervous system, it presumably might not be a function of the *degree* of stimulus isolation. Conceivably this is an all-or-none reaction of the type talked about originally by Lashley (1938).

In addition, the results do not support the hypothesis that providing prior training relevant to an isolated stimulus will cause it to function more efficiently in a paired-associate learning task. Indeed, the weight of data suggests that such relevent pretraining is insufficient to influence attention in the von Restorff effect.

An interesting finding of secondary importance was derived from the analysis of total correct list anticipations. This analysis showed an expected significant effect for improvement over trial blocks but failed to produce a significant effect for isolation procedures. Wallace (1965) reports a series of studies (e.g., Jensen, 1962; Newman & Saltz, 1958) which have consistently demonstrated that lists containing an isolated item are not learned significantly better than control lists which contain no such items. The authors of these studies have concluded that the advantage accruing to an isolated item is seeminly gained at the expense of performance on other items in the list. Wallace (1965) concludes that there has been no consistent finding dealing either directly or indirectly with the von Restorff phenomenon which would suggest that isolation enhances overall list learning. The present findings support this observation and point out a further objection to any practical use of the isolation effect, if total learning is desired for the material into which a heterogeneous term is inserted.

A final observation should be noted regarding subjects who efficiently learned the paired associates—regardless of treatment. The learning task employed in this study was a difficult one as a glance at mean total correct anticipations (Table 5) will show. However, a wide variance in abilities was evidenced by the subjects. Some subjects achieved a 100% level of correct responding midway through the third trial block; some subjects never reliably acquired a single correct association. No data were found to suggest the nature of these individual differences. Nevertheless, it was noted that those subjects who reported the most sophisticated learning strategies and mediational schemes were among the most efficient learners. Generally this involved keying on a single distinct aspect of the Greek stimulus (the individual range of distinct aspects was quite large) and then allowing this aspect to serve as the functional cue which could guide the correct response.

Learners who functioned at the lowest efficiency generally attempted to hook up the global stimulus pattern with the correct response. This suggests that instruction in efficient strategies may be an important variable in facilitating the learning of such associations.

REFERENCES

ANDERSON, R. C., & FAUST, G. W. The effects of strong formal prompts in programmed instruction. *American Educational Research Journal*, 1967, **4,** 345–352.

ANDERSON, R. C., FAUST, G. W., & RODERICK, M. C. Overprompting in programmed instruction. *Journal of Educational Psychology,* 1968, **59,** 88–93.

DUELL, O. K. An analysis of prompting procedures for teaching a sight vocabulary. *American Educational Research Journal,* 1968, **5,** 675–686.

JENSEN, A. R. The von Restorff isolation effect with minimal response learning. *Journal of Experimental Psychology,* 1962, **64,** 123–125.

JONES, J. K. Colour as an aid to visual perception in early reading. *British Journal of Educational Psychology,* 1965, **35,** 21–27.

KJELDERGAARD, P. M., & FRANKENSTEIN, R. *Grapheme-phoneme regularity and its effects on early learning: A pilot study.* Pittsburgh: University of Pittsburgh Learning Research and Development Center, 1967.

LASHLEY, K. S. The mechanism of vision: XV Preliminary studies of the rat's capacity for detail vision. *Journal of General Psychology,* 1938, **18,** 123–193.

NEWMAN, S. E., & SALTZ, E. Isolation effects: Stimulus and response generalization as explanatory concepts. *Journal of Experimental Psychology,* 1958, **55,** 467–472.

OTTO, W. Intralist similarity, order of presentation and color in children's paired-associate learning. *Psychonomic Science,* 1967, **9,** 531–532.

OTTO, W. Elementary pupils' use of cues in paired-associate learning. *Psychology in the Schools,* 1968, **5,** 178–182.

PILLSBURY, W. B., & RAUSH, H. L. An extension of the Kohler-Restorff inhibition phenomenon. *American Journal of Psychology,* 1943, **56,** 293–298.

SAMUELS, S. J. Relationship between formal intralist similarity and the von Restorff phenomenon. *Journal of Educational Psychology,* 1968, **59,** 432–437.

TABER, J. I., & GLASER, R. An exploratory evaluation of a discriminative transfer learning program using literal prompts. *Journal of Educational Research,* 1962, **55,** 508–512.

WALLACE, W. P. Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin,* 1965, **63,** 410–424.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received May 31, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 451-455

EFFECT OF FEEDBACK EXPECTANCY ON CHOICE OF TEACHING STYLES[1]

EDMUND T. EMMER,[2] THOMAS L. GOOD,[3] AND THOMAS D. OAKLAND
University of Texas at Austin

This study investigated whether choice of a teaching style is affected by different types of feedback expectancies. Eighty-seven undergraduate educational psychology students were randomly assigned to four different feedback groups. All students taught a lesson (in a microteaching laboratory), utilizing either a discovery or an expository teaching style. However, each of the four groups received a different expectation of feedback. These were (*a*) how well the student's teaching behavior matched that of his chosen style; (*b*) pupil interest in the lesson; (*c*) pupil learning; (*d*) no feedback. In Group A, preference for an expository style increased after feedback expectancy was induced, suggesting that the expected type of feedback influenced choice of teaching styles. Preference was also related to perceived understanding of the styles and their effects on students interest and learning.

In recent years, microteaching (Allen & Ryan, 1969) has become a well-known technique in teacher education. This approach permits teacher trainees to practice different treaching skills and styles in a variety of real and simulated classroom settings. One advantage of microteaching is that teachers can practice behaviors and learn new or unfamiliar teaching styles without fear of harming or interfering with actual pupils' achievement. Thus, the setting provides an opportunity for maximum experimentation on the part of the trainee with no fear of the consequences of inadequate teaching performance on his pupils.

Another useful feature of the microteaching setting is the opportunity for feedback from peers or supervisors. Feedback can take many forms. For example, it can focus on the effects of the teacher's behavior on the pupils, including the extent to which pupils learned from the lesson or reached whatever objectives were initially intended. Similarly, feedback can be directed at the extent of pupil interest in the lesson or toward the effects of the teacher's behavior upon student motivation for attaining the lesson's objectives. Another potential feedback focus is the adequacy or appropriateness of the teacher's behavior for the particular teaching skill. In other words, if the purpose of a teaching session is to produce a particular pattern of questions (e.g., from narrow to broad) or to practice a particular teaching style (e.g., guided discovery of a principle), feedback can be directed at the extent to which the teacher's behavior approximates some criterion or matches that of some model.

These types of feedback may have differential effects on the behaviors a teacher chooses to practice and may ultimately affect his students' achievement, since other studies have indicated that feedback to the teacher does affect classroom behavior. For example, Tuckman and Oliver (1968) found that pupil feedback to teachers via ratings produced more positive change in teacher behavior than supervisor feedback or no feedback. Wittrock (1962) found that when student teachers expected feedback and evaluation to be based on student achievement, greater gains were made in pupil achievement compared to a control (no achievement set) condition. Thus, it is important to consider in what way the type of feedback provided to the teacher will affect the teaching styles he chooses to utilize. Feedback which resulted in the teacher's choosing only the most familiar or comfortable teaching behavior, rather than to extend his repertoire of instructional skills, would be disappointing in its effect, particularly if the teacher chose a style which he believed was less effective in reaching his instructional objectives.

---

[1] This investigation was supported by Research and Development Center for Teacher Education, University of Texas at Austin, United States Office of Education Contract OE 6-10-108.

[2] Requests for reprints should be sent to Edmund T. Emmer, Department of Educational Psychology, University of Texas at Austin, Austin, Texas 78712.

[3] Now at the University of Missouri.

The present study investigated the influence of the feedback the teacher expects to receive subsequent to teaching on his preference for expository or discovery teaching styles. Four feedback conditions were utilized: (*a*) Appropriateness condition—feedback concerning the extent to which the teacher's behavior matched that of the style (expository or discovery) he chose to practice; (*b*) Learning condition—feedback focusing upon the extent to which students learned from the lesson; (*c*) Interest condition—feedback focusing upon students' interest and motivation in the lesson; (*d*) No-feedback condition—no feedback was provided.

METHOD

The subjects were 87 students enrolled in an undergraduate educational psychology course, the first in a sequence of courses required for secondary teacher certification at The University of Texas at Austin. This study was conducted while the subjects were studying a unit on expository and discovery styles of teaching. In this unit the students read several articles on the topic and taught a lesson to their peers in a microteaching laboratory while utilizing one of the two teaching styles.

At the beginning of the unit, each student completed a three-part questionnaire, the Teaching Style Inventory (TSI), requiring approximately 15 minutes to complete. Part one of the TSI assessed preference for expository and discovery teaching styles. The students were asked to "...assume that you have been asked to teach a group of students for 15 minutes. You are free to pick the subject and particular content to be taught. Which of the two teaching models would you use?" Preference was indicated by circling one of five responses (strongly agree, agree, neutral, disagree, and strongly disagree) to each of two items: "I would use an expository teaching style," and "I would use a discovery teaching style." Responses to each item were scored from one to five, and the difference between the two item's scores was used as a measure of the strength of a subject's preference for either a discovery or an expository style. Thus, negative differences indicate varying degrees of preference for a discovery style over an expository style, zero differences indicate no preference, and positive differences reflect preference for an expository style.

Part two of the TSI was an 11-item checklist of statements concerning the two styles. The subjects indicated whether the particular statement applied to either or both of the two teaching styles. These items related to four possible reasons for preferring either style: (*a*) amount of preparation and instruction time, (*b*) how well the student felt he understood the style, (*c*) effectiveness of the style in reaching learning objectives, and (*d*) effectiveness of the style for developing and maintaining interest and motivation.

Part three of the TSI was a 16-item multiple choice test used to assess understanding of the two styles.

Subsequent to the initial administration of the TSI each student was given the readings that accompanied the unit and told that upon completion of the readings he could sign up for the microteaching part of the unit. Upon returning to sign up, he was readministered the TSI. However, prior to completing the TSI, the experimental condition was induced in the following manner.

Attached to and preceding the TSI were procedural instructions about the microteaching session followed by a description of the type of feedback that was to be made available to the subject about the lesson. An example of these instructions, in this case for Condition B (learning feedback) is provided below.

> *Feedback*
>
> In order that this microteach helps you as a teacher, feedback will be provided to you about your lesson.
>
> This feedback will be made available by asking the students you teach to judge how well they understood the content of the lesson you taught. (When you are a student in your colleagues' lessons, you can also provide the same feedback to them.)
>
> For convenience, a rating form will be used to provide the feedback. This rating scale will be "How adequate was my understanding of the lesson, i.e., how well did I learn what was supposed to be learned?"
>
> / / / / / /
>
> Very Good Understanding — Fair Understanding — Very Poor Understanding
>
> In addition, you will be provided with written feedback, at a later date, made by a trained rater who will listen to the tape recording of your lesson. This feedback will focus upon how appropriately the lesson provided for student learning and the extent to which students appeared to understand and learn from the lesson.
>
> This feedback is not intended as an evaluation; it is simply a means of providing you with information which is relevant to teaching and can aid your development as a teacher.

In the other two feedback conditions (A and C) the instructions paralleled those cited above for the learning condition. In the no-feedback condition, no instructions were given regarding feedback.

The TSIs with the attached feedback descriptions were presented randomly to students when they signed up for microteaching. (However, several additional appropriateness feedback descriptions were inadvertently included, thus producing a slightly larger n for that condition). Students were thus assigned to one of four groups, three of which expected different types of feedback and a fourth group having no expectation for feedback.

After induction of the different feedback expectations, subjects completed the TSI, and then signed up for a time for the microteaching part of the unit. After teaching the lessons, the subjects' preferences for the stlyes again were assessed, using Part one of the TSI.

RESULTS

The results presented below center on two questions: (*a*) What effect did differing expectations of feedback have on preference for a teaching style? and, (*b*) What variables other than feedback expectancy appear related to preference?

An analysis of covariance was computed on the preference scores obtained after inducing feedback expectancy. Initial preference scores were used as the covariate.[4] The differences among the mean preferences for the four groups were significant ($F = 3.00$, $df = 3/82$, $p < .05$). Table 1 presents the initial mean preference scores and the postadjusted mean preference scores.

It is evident from the adjusted means that preference for the teaching styles was altered most in Group A, whose members expected feedback about the appropriateness of their teaching behavior for the style they used. This group's preferences shifted in favor of the expository style. The other two feedback conditions and the no-feedback condition did not differentially affect preferences, although there appears to be a tendency to prefer an expository style more after feedback expectancy than before.

A second analysis of covariance was computed using the preference scores obtained for the subjects after teaching. Again the initial preferences were used as the covariate. The differences among the group means were not significant ($p > .05$), although the relative ordering of the adjusted mean preferences scores remained the same as before teaching (adjusted group means were $\bar{X}_A = .47$, $\bar{X}_B = .24$, $\bar{X}_C = .06$, $\bar{X}_D = -.01$).

TABLE 1

PRETREATMEAT AND POSTADJUSTED MEAN PREFERENCE SCORES

Feedback condition	n	Means of preferences	
		Pretreatment	Post-adjusted
A (Appropriateness)	25	−.40	.97
B (Pupil learning)	21	−.42	.13
C (Pupil interest)	21	−.48	−.14
D (No feedback)	20	.30	.00

Note.—Negative mean preference scores indicate preference for a discovery style.

Reasons for Preference

Data also were obtained on the TSI regarding possible reasons for preferring a style. The subjects indicated whether each of 11 items was true for expository, for discovery, or for both. The items dealt with four variables that might influence preference: how well the teacher felt he understood the style (five items), the styles' effects on student interest (two items), the styles' effects on student learning (two items), and the amount of time required for preparation and instruction (two items). Scores were obtained for each variable by subtracting the number of items checked for a discovery style from the number of items checked for an expository style. Thus, a positive score for a variable means that the subject rated a discovery style higher than an expository style on that characteristic. Table 2 shows mean scores on each variable for three classifications of subjects (preferred expository, no preference, and preferred discovery) for the second administration of the TSI. Table 3 presents the ANOVA for the three preference groups on each variable.

It is apparent that subjects preferring a particular style felt they understood that style better than the style they did not choose. In addition, those who preferred a style were likely to agree that it was more

---

[4] A test of the difference among the group regression slopes was not significant ($p > .25$).

TABLE 2

MEANS OF REASONS FOR PREFERRING A STYLE

Group	n	Reasons for preference			
		Better understood	Student interest	Student learning	Amount of time
Preferred expository	38	2.12	.00	.51	.38
No preference	16	.93	−.63	.06	.25
Preferred discovery	26	−.77	−1.62	−1.16	.05

effective in reaching learning objectives. Those preferring a discovery style felt it was more effective in eliciting pupil interest than those preferring an expository style (however, those preferring an expository style considered both styles equally effective in this regard). There were no significant differences among the preference groups in perceived amount of instruction and preparation time required for the two styles.

TABLE 3

ANALYSES OF VARIANCE COMPARING PREFERENCE GROUPS ON EACH OF FOUR REASONS FOR PREFERRING A STYLE

Variable	MS_B	MS_W	F
Better understood	65.49	3.10	21.12*
Student interest	20.36	1.46	13.94*
Student learning	21.97	1.36	16.15*
Amount of time	.89	1.99	.45

* $p < .001$, $df = 2/77$.

Note.—Abbreviations: MS_B = mean square between groups, MS_W = mean square within groups.

To gain insight into reasons for preference shifts to the expository style in Group A (appropriateness feedback), the responses on the TSI for the subjects in Group A whose preferences changed to expository were compared to the responses of subjects in the other groups whose preferences did not change (but whose initial preferences were the same as the "change" subjects in Group A). Table 4 presents the responses obtained from these subjects before and after feedback expectancy.

The only differential change in the reasons for preference was on the better understood dimension, indicating that subjects who changed their preference to expository also changed their perception of how well they understood the styles. Prior to feedback expectancy these subjects felt they understood the discovery style better than expository, whereas after they learned the focus for feedback in their lesson, they felt, on the average, slightly more understanding for the expository style.

Part three of the TSI, a 16-item multiple choice test of knowledge of the two styles, produced no significant differences among any of the treatment groups.

DISCUSSION

The main finding from this study is that preference for a teaching style can be influenced by the type of feedback the teacher expects to receive. Specifically, in a microteaching laboratory, when the teachers expected to receive feedback based upon the appropriateness of their behavior for the style they chose to practice, they tended to

TABLE 4

MEANS OF REASONS FOR PREFERENCE OBTAINED BEFORE AND AFTER FEEDBACK EXPECTANCY IN GROUP A AND OTHER GROUPS

Reason for preference	Group A: Changed preference to expository (n = 11)			Other groups: Did not change preference to expository (n = 25)			Statistics		
	Before	After	d_1	Before	After	d_2	$d_1 - d_2$	$SD_{d_1-d_2}$	t
Better understood	−1.18	.36	1.54	−1.12	−.92	.20	1.34	.627	2.12*
Student interest	−1.73	−1.64	.09	−1.52	−1.48	.04	.05	.314	.16
Student learning	−1.64	−1.18	.45	−1.20	−.96	.24	.21	.376	.56
Amount of time	−.28	−.28	.00	−.04	.52	.56	−.56	.473	−1.18

* $p < .05$.

abandon a discovery style and shift their preferences toward an expository style.

This shift in preference can be explained, in part, by these teachers' tendency to change their estimate of understanding of the styles. Prior to feedback expectancy a discovery style was thought to be better understood, whereas after feedback expectancy the teachers considered their understanding of an expository style greater than or equal to a discovery style. Teachers with the same initial preferences in the other feedback groups did not change their self-assessment of understanding of the styles. Thus, appropriateness feedback appears to increase the saliency of the degree of understanding of a style. In a situation where performance of a style is open to scrutiny, familiarity may become a dominant characteristic affecting the teacher's decision to use a style. Evidently, the perceived effectiveness of a style for eliciting student interest or student learning is relatively uninfluential since a discovery style was judged superior to expository on these characteristics.

An attempt must be made to explain the failure of the other types of feedback to alter preferences, particularly for the student interest feedback group. As previously indicated, teachers in this group were to receive information about the extent of student interest in and motivation for the content of their lesson. Those preferring an expository style considered it, on the average, to be no more effective than discovery for eliciting student interest; approximately one-half of the group preferring the expository style actually considered discovery a superior style for this purpose. Perhaps the most suitable explanation for their not changing to discovery is that inexperienced teachers' actual concerns are more likely to be focused on themselves, rather than on their students (Fuller, 1969). Thus, the teachers' expectation that feedback will center on student interest is insufficient to overcome the discomfort of trying to alter their tendency to use a more familiar style.

Expectation of feedback can influence preference for a teaching style. However, expected feedback about appropriate use of a style can cause a shift toward a more familiar set of behaviors, even though the style is perceived as less adequate for student learning and student interest. Thus, it appears that the expectation of this type of feedback in a teaching laboratory (or even early in student teaching, when teachers' concerns also may be self-centered) can be detrimental to attempts to encourage teachers to experiment with less familiar patterns of instruction.

REFERENCES

ALLEN, D., & RYAN, K. *Microteaching*. Reading, Mass.: Addison-Wesley, 1969.

FULLER, F. F. Concerns of teachers: A developmental conceptualization. *American Educational Research Journal*, 1969, **6**, 207–226.

TUCKMAN, B. W., & OLIVER, W. F. Effectiveness of feedback to teachers as a function of source. *Journal of Educational Psychology*, 1968, **59**, 297–301.

WITTROCK, M. C. Set applied to student teaching. *Journal of Educational Psychology*, 1962, **53**, 175–180.

(Received June 12, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 456–462

REEXAMINATION OF TWO LEARNING STYLE STUDIES IN THE LIGHT OF THE COGNITIVE PROCESS THEORY OF CURIOSITY[1]

DAVID G. BESWICK[2] AND G. KASTEN TALLMADGE[3]
American Institutes for Research in the Behavioral Sciences, Palo Alto, California

The failure of Tallmadge and Shearer to confirm learning style effects they reported earlier prompted an attempt to reconstruct the two experiments in terms of curiosity theory. Rankings of experimental treatments on eight characteristics theoretically relevant to curiosity arousal were intercorrelated and factor analyzed yielding two arousal characteristic dimensions. When treatment groups were rank ordered on these dimensions, previously unexplained correlations of individual difference measures with performance conformed to a predicted pattern. Tentative questionnaire measures of curiosity were developed from item response data to the Kuder Vocational Preference Record and a biographical inventory. The magnitude of correlations between curiosity measures and performance corresponded to the rankings of treatments on the curiosity arousal dimensions.

Two recent research reports (Tallmadge and Shearer, 1969, 1971) describe apparently conflicting findings. The earlier publication reported statistically significant learner characteristic by instructional treatment interactions (ATI's or learning style effects). The second study also found one significant learning style effect but failed to verify any of the interactions of the earlier research, even where identical individual difference measures and apparently identical instructional treatments were employed. The research reported here represents an attempt to reconcile these disparate findings through reanalysis of the original data in the light of Beswick's (1964, 1970) cognitive process theory of curiosity.

Beswick's theory assumes that each individual has a conceptual "system" for encoding and organizing the stimuli he experiences. Stimuli which cannot be encoded in the existing system produce conceptual conflict. The trait of curiosity is defined as an individual's readiness or predisposition to seek, maintain, and resolve conceptual conflicts.

While curiosity is a characteristic of the individual, it manifests itself only in response to specific situations. Situations which produce conceptual conflict will act so as to arouse curiosity or approach behavior in subjects high in curiosity. They are likely to produce avoidance behavior in subjects low in curiosity. It is assumed that persons who have a high degree of the general trait or characteristic of curiosity will have their learning facilitated by conditions which increase conceptual conflict so that the correlation of curiosity as a trait with learning performance will be a function of the level of arousal due to the introduction or generation of conceptual conflict.

The characteristics of situations which are relevant to curiosity are those most likely to be instrumental in generating conceptual conflict. It is assumed that conceptual conflict occurs when it is difficult for a subject to encode stimulus information into a category system he already possesses. The greater the modification of the previously existing category system which would be required for a signal to be coded, the greater the conflict. Little conflict will arise through the presentation of very diverse stimuli, lacking in structure and unrelated to previous experience, due to the lack of any category system requiring modification. Similarly, when new information is presented in a structure or organization which assists the subject to carry out the coding operation, conflict will be minimal. On the other hand, conflict may be increased either by the introduction of structure or by the introduction of novel stimuli depending on what was lacking in the previous state of affairs.

---

[1] This report is based on research conducted under Contract N61339-68-C-0271 with the Naval Training Device Center, Orlando, Florida, and monitored by G. S. Micheli. The opinions or conclusions expressed herein are those of the authors and should not be construed as necessarily reflecting the views or endorsement of the Navy Department or the Naval Training Device Center. The authors gratefully acknowledge the assistance of T. Whitford (Stanford Center for Research and Development in Teaching), J. W. Shearer, M. P. Rose, and B. M. Lauritz for their assistance in various phases of the work.

[2] Now with the Australian National University.

[3] Requests for reprints should be sent to G. Kasten Tallmadge, Instructional Methods Program, American Institutes for Research, P. O. Box 1113, Palo Alto, California 94302.

Review of the Tallmadge and Shearer (1969, 1971) study findings indicated the possibility that some of the effects observed could have been due to the operation of curiosity. In particular, the Kuder Scientific Interest and Set Toward Education measures appeared to be positively correlated with individual differences in curiosity, and these measures were found to be differently related to performance under different treatment conditions. At the same time, the differences in treatment conditions brought about by the manipulation of instructional methods on the inductive-deductive dimension when combined with differences in subject matters could have produced varying degrees of curiosity arousal due to the generation of different amounts of conceptual conflict. An attempt was made, therefore, to describe the treatments in terms which would enable prediction of the relative magnitudes of correlations between curiosity and performance.

The general strategy was to attempt to discover the moderator variable effects of situational or treatment conditions on the relationship between individual differences in curiosity and criterion performance, or learning, even though the individual differences in curiosity had not been measured directly and relevant characteristics of situations had not necessarily been manipulated by the operations which were employed with other purposes in mind. Since no new data were collected, measurement had to depend on whatever rescaling of data already in hand could be accomplished. This rescaling involved the development of learner curiosity measures from Kuder and Biographical Inventory items as well as the identification and measurement of new treatment dimensions.

METHOD

Scaling of Curiosity Arousal Characteristics of Treatment Conditions

The experimental treatments employed in the 1969 and 1971 studies are described in detail in the cited reports. Each treatment was administered to separate groups of about 58 subjects. Both the treatment conditions and the groups of subjects will be referred to here by the following abbreviations:

TTI—Transportation Technique, inductive method (1969)

TTD—Transportation Technique, deductive method (1969)

ARI 69—Aircraft Recognition, inductive method (1969)

ARD 69—Aircraft Recognition, deductive method (1969)

CNUI—Celestial Navigation, understanding learning course, inductive method (1971)

CNUD—Celestial Navigation, understanding learning course, deductive method (1971)

CNRI—Celestial Navigation, rote learning course, inductive method (1971)

CNRD—Celestial Navigation, rote learning course, deductive method (1971)

ARI 71—Aircraft Recognition, inductive method (1971)

ARD 70—Aircraft Recognition, deductive method (1971).

Reexamination of the treatment conditions as arousal situations for curiosity required that attention be given to some differences between treatments which had not been deliberately manipulated. For this reason, the 10 instructional treatments were rank ordered by Tallmadge and Shearer on each of the following eight characteristics considered by Beswick to be relevant to the arousal of curiosity:

1. Amount of information presented, or the diversity of stimuli
2. Amount of given structure in the presentation of the information, or the assistance given to the subject to organize input
3. Emphasis on output or production, assistance with the structuring of output, or practice in performance
4. Diversity of patterns of organization of input appropriate to the task versus there being only one way to organize the material
5. Extent to which examples were presented before principles versus principles being presented before examples
6. Instructor's tendency to reward attention to the unusual
7. Instructor's tendency to reward attention to general principles
8. Novelty of the situation or task

The two judges who independently ranked the treatment conditions in terms of these characteristics had been responsible for preparing the courses of instruction and had acted as instructors. Both were familiar with the method and content of all the experimental courses. Interjudge reliability of rankings was acceptable on all scales except No. 3, where there was no agreement between judges. For the remaining seven scales, the correlation between assigned ranks was .81. In spite of the lack of agreement on Scale No. 3, a pooled rank order was established for each of the eight scales or dimensions.

Development of Curiosity Measures

Some of the individual difference measures used in the 1969 and 1971 studies appeared to be related to curiosity. Individual item responses to the Kuder Vocational Preference Record and the Biographical Inventory employed in the 1971 study had been tabulated, however, and they were used to develop direct measures of curiosity. Each item was examined independently by three judges for the presence or absence of curiosity imagery as defined by the Curiosity Thematic Apperception Test (Beswick, 1964). The three judges agreed on the scoring of responses for 33 of the items. There was some lack of agreement on 49 items of which 27 were scored by one judge as containing evidence of a minor category of curiosity imagery. After discussion among the judges, 31 items were identified upon which there was a "confident consensus." An additional 25 items were identified for which there was a general consensus but less confidence.

The Kuder and Biographical Inventory items are not binary but multiple choice in nature. For purposes of item analysis, each of the items was regarded as two items—one a positive form and one a negative form. For the positive form of an item, the response categories which scorers agreed contained evidence of high curiosity were given a score of one and the remaining categories, a score of zero. For the negative form, response categories indicating low curiosity were given a score of one and the others, zero. Some response categories were given a score of zero for both the positive and negative forms.

An item-analysis procedure was carried out by computing point-biserial correlations between both forms of each item and total scores. The Kuder items which had been included "with confidence" were found to have very little more internal consistency than those which were considered more doubtful. Similarly, it was found that the positive and negative forms of the items had approximately equal internal consistency and correlations with other variables. In the light of this evidence all Kuder items were combined to form a single scale and all biographical items were combined to form another. The total score obtained from the 56 Kuder items (including both positive and negative forms) will be referred to as Kuder Curiosity, and the sum of the scores on the 16 biographical items as Biographical Curiosity. It did not appear reasonable to combine the two sets into one curiosity scale because of the low correlation between Kuder Curiosity and Biographical Curiosity. This correlation ($r = .33$), while highly significant statistically, nevertheless indicated the possibility of important differences in correlations with other variables.

The degree of internal consistency as indicated by Kuder Richardson Formula 20 reliability coefficients was .83 and .75, respectively. In order not to capitalise on sampling errors, no item selection was carried out. In each scale there were several items which were inconsistent with the others, but these were not eliminated at this stage because no new test data could be collected to cross validate the scale.

Correlations were calculated relating scores derived from the two curiosity scales to other measures including the Kuder Scientific Interest and Set Toward Education scales. (Because of incomplete data, several subjects were eliminated from the item analysis in which data from 339 subjects were employed. Consequently, correlations reported here differ slightly from those reported previously.)

RESULTS AND DISCUSSION

Curiosity Arousal Characteristics of Instructional Treatments

Scores on the eight ranking scales were intercorrelated across the 10 instructional treatments. The obtained correlation coefficients are presented in Table 1.

It is clear from inspection of Table 1 that Scales 1, 4, and 7 produced very similar rank orders and can be interpreted as defining a single dimension. Similarly, ranking Scales 2 and 5 define opposite poles of another dimension. A principal components factor analysis confirmed this interpretation. There were two factors with eigenvalues greater than 1.0 which together accounted for 82% of the total original variance. Factor loadings are given in Table 2.

The two factors were rotated to maximise the loadings of Variables 2 and 5 on the second factor so as to give the best definition of intuitively reasonable dimensions. A Varimax solution produced very similar results. Two orthogonal rank orders of experimental treatments were then defined from the factor scores on the rotated factors. The rank orders of treatment conditions on the dimensions defined by factor analysis is illustrated in Figure 1.

TABLE 1

INTERCORRELATION OF RANKING SCALES FOR CHARACTERISTICS OF EXPERIMENTAL TREATMENTS

Scale	1	2	3	4	5	6	7	8
1	1.00	.63	−.54	.93	−.33	.79	.98	.78
2	.63	1.00	−.10	.43	−.90	.37	.64	.48
3	−.54	−.10	1.00	−.58	−.09	−.28	−.56	.02
4	.93	.43	−.58	1.00	−.17	.71	.94	.74
5	−.33	−.90	−.09	−.17	1.00	−.25	−.41	−.31
6	.79	.37	−.28	.71	−.25	1.00	.71	.76
7	.98	.64	−.56	.94	−.41	.71	1.00	.72
8	.78	.48	.02	.74	−.31	.76	.72	1.00

The interpretation of these two factors may be a matter of some debate. The scales with the highest loadings on Factor I (Scales 1, 4, and 7) indicate a treatment condition in which there is a large amount of information presented which can be organized in many ways while the teacher tends to reward attention to general principles by which such an organization might be achieved. Inspection of the rank order of treatments with the highest factor scores indicates that these treatments are those requiring the greatest attention to abstract problems, a more theoretical understanding, a greater intellectual interest. It is clearly a subject matter dimension and might be termed "ideational versus substantive." The second factor is more closely related to the difference between inductive and deductive teaching methods. The two ranking scales which best define the factor are concerned with the extent to which the structure or organization of the material is made explicit by the instructor. At one extreme we have treatment conditions in which the learner must discover structure for himself. At the other extreme, a prescribed organization is given in a relatively didactic manner. It might be termed "implicitness versus explicitness" of the organization of subject matter.

TABLE 2

FACTOR ANALYSIS OF CHARACTERISTICS OF EXPERIMENTAL TREATMENTS

Treatment characteristics	Factor loadings				h^2
	Principal component		Rotated factor		
	I	II	I'	II'	
1	.98	.15	.87	.47	.98
2	.71	−.63	.19	.93	.90
3	−.46	−.61	−.74	.22	.59
4	.91	.32	.92	.29	.94
5	−.48	.81	.10	−.94	.89
6	.81	.13	.72	.38	.67
7	.98	.09	.84	.51	.96
8	.81	.09	.59	.56	.66
Proportion of trace	.63	.20	.47	.35	.82

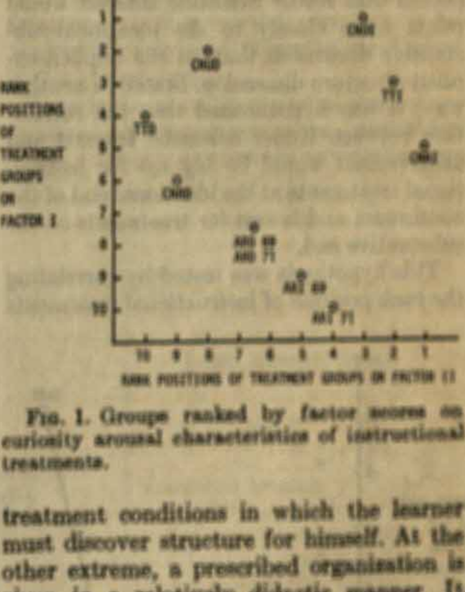


FIG. 1. Groups ranked by *factor scores* on *curiosity arousal characteristics of* instructional treatments.

Learning Style Effects of Curiosity

It had been expected that the Kuder Scientific Interest scale and the Set Toward Education measure would be positively related to curiosity. To explore this possibility a set of responses to the Kuder and the Set Toward Education were generated for a hypothetical subject with very high curiosity. It was found that such an imaginary person would score at the 97th percentile on the Kuder Scientific Interest scale, and more than two standard deviations above the sample mean on the Set Toward Education scale. It should also be noted that in spite of the assessment that a highly curious person would score highly on both scales there is very little correlation ($r = .08$) between the two scales. Because of this low correlation between the scales, it might be expected that the two measures are related to different aspects of curiosity.

It also seemed likely that the two scales would be related to the different aspects of the curiosity arousal characteristics of instructional treatments revealed by the factor analysis described above. It was expected that Kuder Scientific Interest would relate more closely to the ideational-substantive dimension than to the implicit-explicit structure dimension. Stated in another way, it was hypothesized that the correlation between Kuder Scientific Interest and achievement would be highest for instructional treatments at the ideational end of the continuum and lowest for treatments at the substantive end.

This hypothesis was tested by correlating the rank position of instructional treatments on Factor I with the rank order of the correlations between Kuder Scientific Interest and performance within the corresponding treatment groups. This correlation was found to be .75, statistically significant at the .05 level.

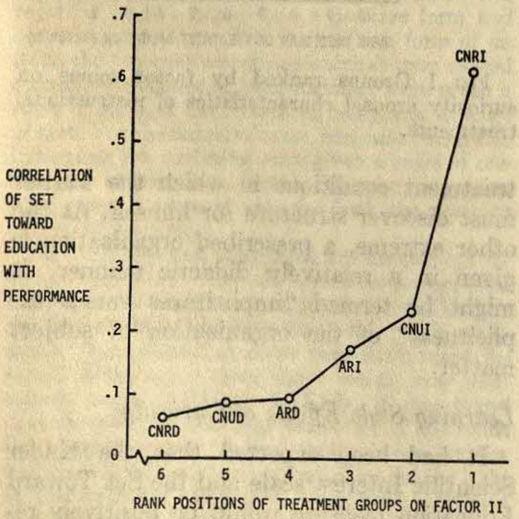


FIG. 2. Relation of within-group correlations between Set Toward Education and performance to ranking of treatments on curiosity arousal Factor II.

Scores on the Set Toward Education scale were, unfortunately, available only for subjects who participated in the 1971 study. At its face value, the scale provides a measure of interest in learning about general principles as opposed to isolated facts. It might seem then that both curiosity arousal dimensions could be related to it. It was predicted, however, that the Factor II type of arousal would be the more relevant since a person who has a strong interest in discovering the organization of material could be expected to be particularly aroused when material to be learned has an implicit rather than an explicit structure. An empirical check of this hypothesis revealed that correlations between Set Toward Education and performance were unrelated to the rank order of treatments on Factor I but, as shown in Figure 2, were highly related to rank order on Factor II. Indeed, it is somewhat embarrassing to report that the correlation between the rank orders was 1.00.

It must be assumed that the perfect correspondence between the two rank orders is merely fortuitous. Nevertheless, it is of some theoretical significance that the previously unexplained high correlation ($r = .62$) between Set Toward Education and performance in the Celestial Navigation Rote Inductive treatment (CNRI) is now made understandable in view of the top ranking of that treatment on Factor II.

The Kuder Curiosity scale, as expected, was found to be highly correlated with Kuder Scientific Interest ($r = .87$, $p < .01$). Kuder Curiosity, however, was considered to be theoretically more basic than the Scientific Interest scale. This expectation was confirmed in the following manner: when Kuder Curiosity was partialed out of the correlation between Scientific Interest and performance the resulting partial correlation was not significantly different from zero ($r = .01$). On the other hand, when Scientific Interest was held constant, the partial correlation of Kuder Curiosity with performance remained significantly greater than zero ($r = .14$, $p < .05$) for the total sample.

When within-treatment-group correlations between Kuder Curiosity and performance were rank ordered and compared to the rank order of the treatments on Factor I, a correlation of .66 was found. Although not statistically significant ($p > .05$, one-tailed), this correlation was of approximately equal magnitude to that observed with Kuder Scientific Interest. Partialing out the effects of intelligence (as measured by the Navy General Classification Test) had the effect of causing the within-group partial correlations to be more closely related to the rank order of treatments on Factor I ($r = .89$, $p < .01$, one-tailed).

Biographical Curiosity appeared to be a different trait from Kuder Curiosity, at least in some respects. The highest correlation of Kuder Curiosity with performance and the only negative correlation of Biographical Curiosity, for example, occurred in the same treatment group (CNUD). Because of this difference, it was hoped that Biographical Curiosity would relate differentially to performance as a function of Factor II treatment characteristics in a manner similar to the Set Toward Education measure. The relationship, however, was a weak one ($r = .43, .20 > p < .10$). The highest correlation did occur in the predicted treatment condition (CNRI), but all the Biographical Curiosity correlations with performance were low and their rank order was not highly related to the scaling of situational characteristics on Factor II. The weakness of this relationship was not surprising in view of the fact that the correlation between Biographical Curiosity and Set Toward Education was only .07.

Despite the differences between the two curiosity measures, they were not entirely independent ($r = .33$). The common element appeared to be an interest in physics and chemistry. Items concerned with experimenting are present in both. Such items in the Biographical Inventory correlated significantly with the Kuder Curiosity scale. There were other items included in the Biographical Curiosity measure which did not correlate with the Kuder Curiosity scale. These were items concerned with reading magazines and newspapers, using telescopes, and belief that one has gained more knowledge from one's own reading and observation than from school and family. Other items showing a stronger relationship with Biographical Curiosity than with Kuder Curiosity were those concerning the use of a library for purposes other than course work and an interest in reading in general.

The question was asked whether the "nonscientific" interest in reading about, rather than manipulating, the world could be related to the effects of arousal situations in a manner different from that which was observed with the more scientifically oriented Kuder Curiosity measure. To address this question, variance due to Kuder Curiosity was partialed out of the correlations between Biographical Curiosity and performance yielding partial correlations between "Biographical Nonscientific Curiosity" and performance.

In order to relate these partial correlations to nonscientific treatment arousal conditions, it was necessary to perform an operation on

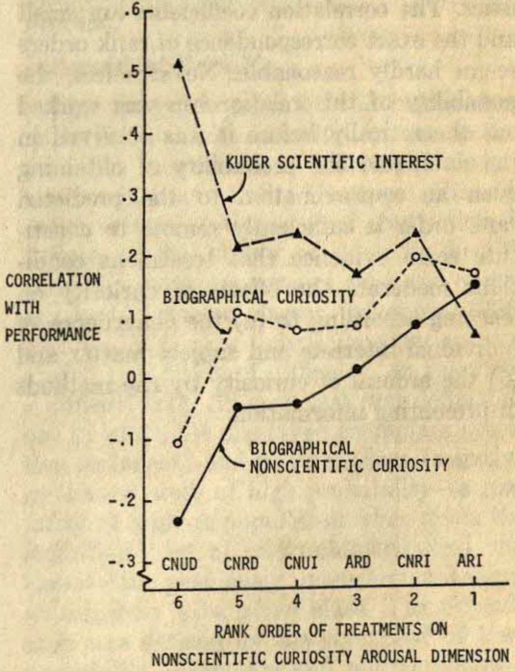


FIG. 3. Correlations with performance related to ranking of treatments for arousal of nonscientific curiosity.

the treatment conditions similar to that performed on the Biographical Curiosity measure. To accomplish this objective, the rank position of each treatment on Factor I was subtracted from its rank position on Factor II. (This procedure was equivalent to creating a new treatment dimension from the upper left corner of Figure 1 to the lower right corner.)

When the rank order of these differences was compared with the rank order of the correlations of Biographical Nonscientific Curiosity with performance, the correspondence was again found to be perfect (r = 1.00). Figure 3 illustrates this relationship as well as those of Kuder Scientific Interest and Biographical Curiosity to the rank differences of treatments on Factors I and II.

In summary it may be stated that when we have primarily or only Factor II type arousal, then that "nonscientific" kind of curiosity which is measured by the Biographical Inventory alone is relatively facilitating of performance compared with the debilitating effect found when arousal is given only by Factor I treatment characteristics. The correlation coefficients are small and the exact correspondence of rank orders seems hardly reasonable. Nevertheless, the possibility of this relationship was worked out theoretically before it was observed in the data, and the probability of obtaining even an approximation to the predicted rank order is sufficiently remote to constitute some evidence that treatment conditions moderate the effects of curiosity on learning according to (*a*) the congruence of individual interests and subject matter and (*b*) the arousal of curiosity by the methods of presenting information.

Results of the several analyses undertaken in this review of earlier learning style studies indicate the advisability of undertaking further research in the area based on curiosity theory. Statistical techniques of manipulation a posteriori cannot substitute for experimental controls and a research design developed from a theoretical model of basic processes. The somewhat unorthodox techniques of analysis employed in this review allow only tentative conclusions from a partial simulation of an experimental study. To the extent that it has been possible to identify the relevant variables, it appears that the previous results can be seen in part as a function of the learning style effects of curiosity. Confirmation must await direct experimental manipulation of conceptual conflict inducing characteristics of instructional treatments and direct measurement of the predisposition to curiosity behavior, together with adequate control over such important covariates as intelligence, anxiety, and subject matter differences.

REFERENCES

BESWICK, D. G. Theory and measurement of human curiosity. Unpublished doctoral dissertation, Harvard University, 1964.

BESWICK, D. G. Cognitive process theory of individual differences in curiosity. Paper presented at the Conference on Intrinsic Motivation, Ontario Institute for Studies in Education, Toronto, June 1970.

TALLMADGE, G. K., & SHEARER, J. W. Relationships among learning styles, instructional methods, and the nature of learning experiences. *Journal of Educational Psychology*, 1969, **60**, 222–230.

TALLMADGE, G. K., & SHEARER, J. W. Interactive relationships among learner characteristics, types of learning, instructional methods, and subject matter variables. *Journal of Educational Psychology*, 1971 **62**, 31–38.

(Received June 22, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 463–467

INTERACTION OF ANXIETY WITH PERFORMANCE ON TWO LEVELS OF TASK DIFFICULTY[1]

ROBERT D. TENNYSON[2] AND F. ROSS WOOLLEY
Brigham Young University

A self-instructional concept acquisition program was divided into difficult and easy sections based on empirically determined probability ratings of the positive instances and negative instances. The prediction of a disordinal interaction was tested using measured anxiety levels as one effect and task difficulty levels as the other effect. Anxiety was measured by the State-Trait Anxiety Inventory (STAI) and blood pressure. The scores for anxiety on the STAI A-State Scale increased significantly for the 29 college students during the difficult task and decreased with the easy task ($p < .01$). When the difficult task and the easy task were compared, a disordinal interaction resulted, that is, the relative mean error performance of the high A-State subjects and the low A-State subjects was reversed ($p < .05$).

Empirical data on the interaction of emotional effects with the cognitive effects of instruction is minimal. In a study by O'Neil, Spielberger, and Hansen (1969) the anxiety state of subjects involved in a self-instructional learning task was significantly related to the difficulty of the subject matter. Their study, with tightly controlled internal validity, was an investigation of Spielberger's (1966) assumption that state anxiety (A-State) differs from trait anxiety (T-State). A-State environmental conditions fluctuate while T-State is the variable related to more stable anxiety in individuals. The study by O'Neil et al. also investigated the Spence (1958)–Taylor (1956) drive theory that high A-State persons would commit more errors on a difficult task than low A-State persons, but this situation would reverse on an easy task. The O'Neil et al. study in part confirmed this assumption when the subjects did increase in anxiety during the difficult ask. The study showed an error reverse interaction within the difficult task.

Cronbach and Snow (1969) discuss the loss of empirical data in instructional research due to the failure of experiments to analyze the effects of interactions. If the relationships between the dependent and independent variables are plotted graphically under varying blocking schemes, the resulting slopes will be different. A disordinal interaction results when the dependent variables significantly reverse their measured means according to the level of the independent variable. Measuring the interaction between the learner's emotional behavior and the treatments is necessary when studying emotional behavior in the instructional process.

Independent Variable

This study investigated the independent variable of task difficulty in connection with the interaction of individual anxiety states and instructional treatment. Classification behavior, the level of instruction, is defined as the ability of an individual to generalize within a concept class and to discriminate between classes (Mechner, 1966; Merrill, 1971). The independent variable included two levels of difficulty, an easy and a difficult task. An easy task was defined as one in which the displayed exemplars (positive instances) and nonexemplars (negative instances) were of high probability—a majority of a given population when given the definition (list of relevant attributes) can classify the previously unencountered items as members of a given class. The difficult task was defined as one composed of low-probability exemplars and nonexemplars—defined as those previously unencountered instances which a majority of a given population when given the definition cannot classify correctly (Woolley & Tennyson, 1971).

---

[1] This research was conducted in the Department of Instructional Research and Development, Brigham Young University.

[2] Requests for reprints should be sent to Robert D. Tennyson who is now at the Department of Educational Research and Testing, Florida State University, Tallahassee, Florida 32306.

Dependent Variable

The dependent variable was the subject's error rate recorded for each level of the subject's A-State. Subjects with a low measured A-State were hypothesized to make fewer errors on the difficult task than subjects with a high measured A-State. Conversely, high A-State subjects were hypothesized to make fewer errors on the easy task than low A-State subjects. Third, a subject's measured A-State score was hypothesized to increase following the difficult task and to decrease with the easy task. Thus, the study hypothesizes that a disordinal interaction exists between task difficulty and the subject's state anxiety.

METHOD

Subjects

The instance probability analysis was conducted with 35 students enrolled in an undergraduate educational psychology class at Brigham Young University. The program sessions involved 29 randomly selected undergraduate general psychology students (12 males and 17 females).

Task

The self-instructional program for this study was concept classification. The concept's relevant attribute was poetic trochaic meter. An instance probability analysis involving 85 pieces of poetry was conducted to determine correct classification probability ratings for each piece of poetry. The subjects studied a definition of trochaic meter and then identified each poetic selection as either an example or not an example. A frequency distribution of correct classifications resulted which resembled a normal one-tailed curve. High-probability selections were those correctly defined by the subjects at a frequency of 60% level and above; low-probability selections were identified at 30% level for exemplars and at 50% level for nonexemplars.

The task was composed of two parts: a difficult section of poetry selection, and an easy section of poetry selections. The definition was not available to subjects for further reference. Responses were recorded on an IBM answer sheet. No feedback was given concerning the correctness of the responses.

The definition of trochaic meter presented to subjects was:

> Part of the rhythm of a poem is determined by the time between stresses being occupied with unstressed syllables or pauses. To denote the stress patterns is to establish the meter. One of the major meter scansions is named trochee and consists of a stressed syllable followed by an unstressed syllable (marked thusly:/◡). Trochaic meter is often employed when the author wishes to convey intense emotion—usually sarcasm, anger, or bitterness. The lines sometimes end with a stressed syllable rather than a weak syllable. This is called a masculine ending. Chants, such as used by American Indians, are often set to trochaic beat.

Anxiety Measures

The Spielberger, Gorsuch, and Lushene (1969) State-Trait Anxiety Inventory was used to measure A-Trait and A-State. The A-Trait (Form X-2) Scale asks subjects to indicate how they "generally feel," while the A-State (Form X-1) Scale requires subjects to indicate how they feel "at this moment." Both scales consist of 20 items and are on opposite sides of the questionnaire. Subjects responded according to a 4-point scale using these statements as the criteria for the A-State Scale: Very much so (4); Moderately so (3); Somewhat (2); and Not at all (1). The four categories on the A-Trait Scale are: Almost always (4); Often (3); Sometimes (2); and Almost never (1). Items are stated in a random sequence of negative and positive scoring schemes to eliminate response sets. Scores range from a minimum (low anxiety) of 20 to a maximum (high anxiety) of 80. Systolic blood pressure (SBP) was taken by the standard procedure using a blood pressure cuff (sphygmomanometer) and stethoscope.

Experimental Design

This study utilized a multiple treatment experimental design (Campbell & Stanley, 1963). The one-day study involved subjects as the experimental unit. Three experimenters worked with a maximum of eight subjects per session. A session consisted of four periods in which the systolic blood pressure and STAI A-State Scale were administered.

During the pretask period each subject took the STAI A-Trait Scale followed by the STAI A-State Scale and the SBP. The subject was then given directions and the definition. The second period was the difficult task followed by the SBP and the STAI A-State Scale. In the third period, the subject again received general directions and proceeded with the easy task followed by the SBP and the STAI A-State Scale. During the fourth period, the subject was alone for 3 minutes. The SBP was taken and the STAI A-State Scale was administered a fourth time.

Procedure

The experiment took place in a room which contained eight individual cubicles. A small desk and two chairs were in each cubicle; one chair for the subject and one for the nurse when taking the SBP. The subject was isolated in the cubicle and remained seated until finished. The experimental time per subject averaged 45 minutes. The experimenters entered the cubicle only to read directions, record blood pressure, and administer the anxiety scales.

During the pretask period the subject was given a direction sheet which explained the appropriate procedures for the program. Each time the subject finished a section he was instructed to open his door to indicate to the experimenter that he was finished. The directions to the STAI A-Trait Scale were read aloud to the subject emphasizing, "How you feel at this moment." When finished, the SBP was taken for the first time.

Immediately following the pretask procedures, the subject was given the definition of trochaic meter which he could study as long as desired but it was taken away during the performance period. The directions were written to produce a condition of mild stress, that is, the directions explained that the task consisted of complex selections of poetry. When the subject opened the door he was given the difficult task. Directions explaining how to mark on the IBM answer sheet were contained in the booklet. The subject was directed to continue with the program until completed. When finished, the SBP was read and the STAI A-State Scale was administered emphasizing, "How you felt while taking the difficult task."

The subject was then given the easy task with instructions designed to produce less stress, that is, the directions explained that the task consisted of simple selections of poetry. When the subject opened the door, his SBP was taken and the STAI A-State Scale was administered for the third time.

The final period began after a 3-minute wait. The SBP was taken for the fourth time and the STAI A-State Scale was given with the instructions to indicate, "How you feel at this moment." The subject left the cubicle when completed and was thanked for his participation.

RESULTS

The dependent variable measures were STAI A-State Scale and A-Trait Scale scores, SBP readings, and task errors. The STAI A-Trait Scale score was correlated .44 with the second (difficult period) STAI A-State Scale score. An *r* of .62 was calculated between STAI A-Trait Scale and the third measure (easy period) of STAI A-State Scale score. Assumptions for linearity of the data were checked by an eta test and homoscedasticity for equal variance. The subjects were separated into relatively low and high anxiety groups according to their score on the first STAI A-State Scale. The median score was 39. There was a correlation of .92 between the subjects initial classification and his classification on the second (difficult) and the third (easy) STAI A-State Scales.

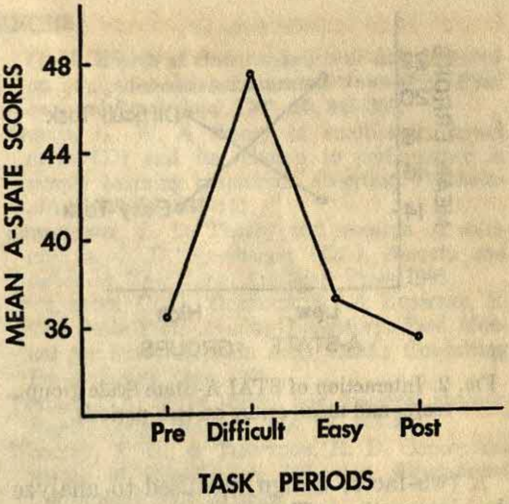


FIG. 1. Mean STAI A-State Scale scores across task periods.

An analysis of variance of repeated measures was used with a Newman-Keuls Sequential Test to obtain significant differences of STAI A-State Scale means. There was a difference between anxiety periods ($F = 6.58$, $df = 3/84$, $p < .01$). Figure 1 represents the mean scores over the four periods for all subjects combined. The STAI A-State Scale scores increased from the pretask period to the difficult task and decreased following the easy task with no change within the posttask period. The means differences (Table 1) were between the difficult period and the other periods ($p < .01$), but there were no significant differences between the pretask and easy task or posttask ($p > .05$). Blood pressure readings did not vary from period to period ($F = 1.07$, $df = 3/112$, $p > .05$).

TABLE 1

NEWMAN-KEULS TEST OF STAI A-STATE MEANS

Period	$\bar{x}$	Post	Pre	Easy
Difficult	46.34	10.66*	10.14*	9.45*
Easy	36.89	1.21	.69	
Pre	36.20	.52		
Post	35.68			

Note.—$N = 29$; $MS = 96.22$.
* $p < .01$.

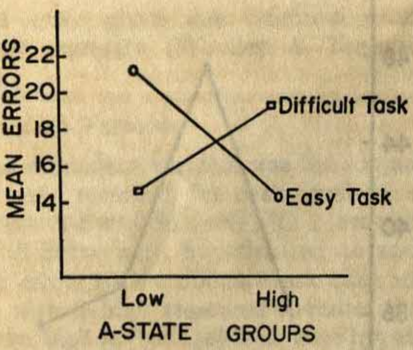


FIG. 2. Interaction of STAI A-State Scale group scores and mean errors on task periods.

A two-factor design was used to analyze the interaction. Task treatment, at two levels, was administered to all subjects, while the anxiety effect was controlled with two different groups. For the main effects there were no differences ($p > .05$) for either the between-subjects on anxiety or the within-subjects on task difficulty; however, the interaction was different ($F = 11.26$, $df = 1/28$, $p < .01$). The error means were plotted according to anxiety groups and task difficulty (Figure 2). To test for differences between the means within anxiety level, the least significant difference test was used because of its decreased probability of making a Type II error (see Table 2). The low A-State group differed between tasks ($p < .05$). A difference was found between the mean errors of the tasks for the high A-State group ($p < .05$). The low A-State group made fewer errors on the difficult task than the high A-State group ($p < .05$). The disordinal interaction was complete with the high A-State group making fewer errors on the easy task than the low A-State group ($p < .05$). To check the results of the least significant difference test an individual degrees of freedom test (Table 2) was included; the results were similar ($p < .05$).

TABLE 2

MEAN ERROR INTERACTION

A-State group	Task periods		Difference	*F*
	Difficult *MS*	Easy *MS*		
Low	14.64	20.85	6.07[a]	10.68[b]
High	19.07	14.78	4.28[a]	5.12[b]

[a] Least significant difference = 3.80 where $p = .05$.

[b] $F = 4.20$, $df = 1/28$, $p = .05$.

DISCUSSION

Results of the STAI A-State Scale scores indicate that the subjects did exhibit an increase in state anxiety when performing the difficult task of low probability exemplars and nonexemplars. These findings support O'Neil et al. (1970). Subjects STAI A-State Scale scores decreased to a level that was not significantly different from the pretask and posttask levels during the easy task period. This corresponds with Spielberger's contention that a person's state anxiety does fluctuate in instructional situations known to be difficult or easy.

The disordinal interaction of measured high-anxiety individuals with low-anxiety individuals on the two levels of task difficulty follows the assumptions of the Spence-Taylor drive theory. The main factor in obtaining a reversal is the control of the independent variable. The tasks were constructed on empirically based data so that the difficult task was known to be more difficult than the easy task. Thus the condition contributing to a possible interaction effect between the A-State Scale scores and task were systematically controlled.

The SBP readings were not indicators of anxiety increases. This is consistent with the O'Neil et al. (1970) findings that SBP correlation with performance was not significant.

Since the data did result in a significant disordinal interaction, several implications for instruction are evident. Individuals who do have anxiety increases during difficult tasks might be expected to perform more efficiently if they receive instruction geared for slower increases of difficulty. The opposite expectation would be indicated for low state-anxiety individuals. For these individuals, to be instructed with an easy task would produce a less effective means of learning. The disordinal interaction of state anxiety and instructional complexity cannot be ignored in instructional systems.

REFERENCES

CAMPELL, D. T., & STANLEY, J. C. *Experimental and quasi-experimental designs for research.* Chicago: Rand-McNally, 1963.

CRONBACH, L. J., & SNOW, E. C. Individual differences in learning ability as a function of instructional variables. (Final rep., Stanford University, Office of Education, United States Department of Health, Education, and Welfare. ETC. BR-6-1269. March, 1969, 221P. Contract-OEC-4-6-061269-1217)

MECHNER, F. Science education and behavioral technology. In R. Glaser (Ed.), *Teaching machines and programed learning, II: Data and directions.* Washington, D. C.: National Education Association, 1966.

MERRILL, M. D. Necessary psychological conditions for defining instructional outcomes. In M. D. Merrill (Ed.), *Instructional design: Readings.* Englewood Cliffs, N. J.: Prentice-Hall, 1971.

O'NEIL, H. F., SPIELBERGER, C. D., & HANSEN, D. H. Effects of state anxiety and task difficulty on computer-assisted learning. *Journal of Educational Psychology*, 1969, **60**, 343–350.

SPENCE, K. W. A theory of emotionally based drive (D) and its relation to performance in simple learning situations. *American Psychologists*, 1958, **13**, 131–141.

SPIELBERGER, C. D. Theory and research on anxiety. In C. D. Spielberger (Ed.), *Anxiety and behavior*. New York: Academic Press, 1966.

SPIELBERGER, C. D., GORSUCH, R., & LUSHENE, R. *The State-Trait Anxiety Inventory, Test Manual for Form X.* Palo Alto, Calif.: Consulting Psychologists Press, 1969.

TAYLOR, J. A. Drive theory and manifest anxiety. *Psychological Bulletin*, 1956, **53**, 303–320.

WOOLLEY, F. R., & TENNYSON, R. D. Conceptual model of classification behavior. *Educational Technology*, 1971, in press.

(Received June 22, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 468–473

DIVERGENT THINKING IN BLIND CHILDREN[1]

WILLIAM J. TISDALL, A. EDWARD BLACKHURST[2]
University of Kentucky
AND CLAUDE H. MARKS
University of Delaware

The purpose of this investigation was to study the influence of visual deprivation upon divergent thinking dimensions of intelligence. Six divergent thinking tests were administered to 76 seeing children, 76 blind children in residential schools, and 76 blind children in day school programs. It was concluded that (*a*) blind children exhibit more verbal fluency than do seeing children; (*b*) visual familiarity with the environment allows seeing children advantage over the blind in a small number of divergent thinking activities; (*c*) blind and seeing children (CA 10–12) generally do not differ in the ability to think divergently; (*d*) blind children in residential and day school settings are equally capable of thinking divergently; and (*e*) seeing and day school blind males tend to be more divergent than their female classmates.

The phenomenon of divergent thinking has received considerable attention in the research literature of education and psychology in recent years. The formulation of Guilford's (1960) theoretical model of the Structure of Intellect has provided researchers with new avenues for more penetrating scrutiny of previously unexplored aspects of intellectual functioning. In the present study, divergent thinking is defined as that kind of thinking in which new information, or new combinations of ideas, are generated out of given or known information and which represents a respondent's performance on purported verbal measures of originality, fluency of ideas, flexibility of thought, and elaboration of ideas. According to Guilford's theory, divergent thinking is a vital prerequisite to creativity which in turn is an important part of intelligence.

The general purpose of this investigation was to study the influence of visual deprivation upon the verbal divergent thinking dimension of intelligence. Specifically, the primary objectives were to study differences which may exist in the divergent thinking abilities of (*a*) blind children and seeing children, (*b*) residential school and day school blind children, (*c*) both blind and seeing males and females, and (*d*) children of varying ages of onset of blindness. For purposes of this investigation, blindness was defined as residual vision of light perception or less (Jones, 1963).

Although some research has been conducted on the creative expressions of blind children Lowenfeld (1952), it was done through the artistic medium of modeling. The need remains for an examination of the creative abilities of blind children along the intellectual, in addition to the artistic, dimension. The effect which blindness has on the development of divergent thinking abilities of blind children is not known. Yet, there is some support for the contention that the blind do possess the ability to think divergently. Guilford (1959), for example, argued that all persons possess the various abilities defined in his theoretical model in differing degrees since the abilities are assumed to be continuously distributed variables. In addition, the work of Wilson, Guilford, and Christensen (1952) led to the conclusion that the intellectual component of originality is a continuous variable

---

[1] The authors gratefully acknowledge the assistance of nearly 50 persons in residential and public school programs for blind and seeing children and colleagues at the Universities of Pittsburgh. Special thanks are due J. P. Guilford and E. Paul Torrance for their permission to use their tests of divergent thinking in this study, which was supported by Grant Number 32-27-0350-6003, Bureau of Education for the Handicapped, United States Office of Education.

[2] Requests for reprints should be sent to A. Edward Blackhurst, UKRSEIMC, 730 South Limestone Street, Lexington, Kentucky 40506.

which is possessed by all individuals to some degree.

Lowenfeld (1959) presented a strong argument for the existence of creative ability in everyone. He maintained that man possesses the ability to create intentionally while the animal does not. Thus, it follows that every man is a potential creator. At the same time, Lowenfeld indicated that the extent to which the creative ability is developed varies among individuals. Each person, he claimed, has functional and potential creative abilities. The former is that which is used by the individual while the latter is that portion of the person's creative ability which remains unused. Some individuals, because of their early training and experiences, have matured into essentially functional creators. Others, not having gained an awareness of their potential, have remained uncreative.

While Guilford's theoretical explanation of the nature of intelligence posits the potentiality of all individuals to think divergently, Lowenfeld implies that deterrents to an awareness of that potential would lead to a truncation of divergent thinking ability. Since the severe sensory impairment of blindness leads to a supression of interaction with the environment, blind individuals may be more likely to develop as potential, rather than functional, divergent thinkers. Seeing children, on the other hand, having use of the visual sense modality, may be more likely to have greater and more varied opportunities to become functional divergent thinkers. The present study was designed to determine if this is the case.

METHOD

Subjects

The sample consisted of 76 seeing children (40 males, 36 females), 76 blind children enrolled in public day schools (39 males, 37 females), and 76 blind children in residential schools for the blind (34 males, 42 females). All subjects were between 10 and 12 years of age and were enrolled in school programs in the Eastern half of the United States. The characteristics of the groups selected for the investigation are reported in Table 1.

For purposes of determining group comparability, analyses were conducted between groups on the independent variables. All analyses were two-tailed *t* tests for differences between means (Guilford, 1956). The .05 level of significance was preset as the criterion for rejecting the null hypotheses.

TABLE 1

MAJOR CHARACTERISTICS OF THE SAMPLE

Variable	Seeing children		Day school blind children		Residential school blind children	
	M	*SD*	*M*	*SD*	*M*	*SD*
CA (in months)	141.2	8.4	140.0	10.0	140.2	9.2
MA (in months)	163.8	18.9	149.0	23.6	149.5	27.2
IQ	116.1	12.9	106.5	14.8	106.3	16.6
Years in school	6.08	0.67	5.93	1.0	5.81	1.15
Grade	6.08	0.67	5.47	1.0	5.17	1.3

It was found that there were no differences between the two blind groups on the independent variables. However, the seeing group was significantly superior to both blind groups in MA and IQ. Even though these differences were found, because of the low correlation reported between intelligence and tests of divergent thinking (Torrance, Yamamoto, Schenitzki, Palamutlu, & Luther, 1960), it was felt that these differences would not differentially affect the divergent thinking test scores. Furthermore, and more important, the intelligence criterion was construed for purposes of this study as a screening device only in order to ascertain that no mentally retarded subjects were included in the sample. In addition, IQ comparisons between the blind and seeing subjects in this study were viewed with caution, since the same tests were not used with all subjects.

It was found that there were significant differences among the three groups with respect to grade level in school. Seeing children tended to be in higher grade levels than both groups of blind children. These differences in grade level could be a function of learning difficulties encountered by the blind children—particularly in the residential schools. This is supported by the fact that years in school corresponded exactly to grade level for the seeing subjects, whereas the blind children were placed in grade levels lower than would be expected on the basis of the number of years they had been in school. This is consistent with the findings reported by Lowenfeld (1963) that blind children are overage for their grade levels.

Instrumentation

Six tests of verbal divergent thinking were administered to each subject. From the 17 subtests, 41 separate divergent thinking scores were derived. The following tests were administered:

1. *Word Fluency* (Christensen & Guilford, 1958) yielded a total fluency score in addition to scores on Subtests B and T.

2. Fluency, Flexibility, Originality, and Elaboration were assessed with the *Product Improvement* test (Torrance & Palamutlu, 1960).

3. The Brick, Heat, and Water Subtests of the *Unusual Uses* (Undated) battery yielded Fluency, Flexibility, and Breadth scores as well as a total Unusual Uses Fluency score.

4. Fluency, Flexibility, Breadth, Remote, and Obvious scores were obtained on each of the Food, Reading, and Writing, and Balance items of the *Consequences Test* (Christensen, Merrifield, & Guilford, 1958). A total Consequences Fluency Score was also derived.

5. The Drink, Sweet, Smooth, and Green Subtests were administered to determine *Ideational Fluency* (Christensen & Guilford, 1957). These also yielded a total Ideational Fluency Score.

6. Scores were also obtained on the Hammer, Wind, and Glue Subtests of the *Seeing Problems Test* (Undated).

A brief pilot study was conducted for the purpose of determining whether or not blind children could respond meaningfully to the tests of divergent thinking which were used in this study. The tests were administered to 10 blind children, between the ages of 10 and 12. It was found that blind subjects were capable of providing meaningful and relevant responses to these verbal tests of divergent thinking.

Methods of Analysis

When scoring the tests of divergent thinking, it was found that the scoring criteria proposed by the authors of the tests were not, in a number of cases, sufficiently broad enough to permit categorization of the subjects' responses. It was therefore necessary to revise the scoring criteria.

Upon revision, the extent of agreement among scorers as to the application of the criteria was determined. The test papers of 15 subjects were randomly selected for each of the six tests of divergent thinking. Three members of the project staff scored each test independently according to the revised scoring criteria. The total number of responses was then computed for each of the tests as was the number of instances in which the scorers did not agree. The percentage of interscorer agreement was then calculated.

In 10 of the 17 subtests the interscorer agreement was 90% or higher. The lowest percentage of agreement was 85.2. Provisions were then made to assure the accuracy and consistency of the scoring. That is, test scorers were instructed to discuss questionable responses which arose during the scoring process and reach agreement before assigning that response to a category.

To guard against scorer bias, all test papers were assigned random numbers prior to scoring by a person not associated with the project. Thus, the scorers were not aware whether the tests which they corrected were those of day school blind, residential school blind, or seeing children.

To test for significant differences between groups, means scores on each of the variables were computed for each group. Comparisons between means were then performed using t tests. The following comparisons were made: (a) seeing versus day school blind subjects, (b) seeing versus residential school blind subjects, and (c) day school blind versus residential school blind subjects. Differences between mean scores of males and females in each of the three groups received similar analysis. The .05 level of significance was set as the criterion for rejecting the null hypotheses. All statistical analyses were performed using raw-score data.

RESULTS AND DISCUSSION

Mean divergent thinking test scores and standard deviations for the three groups are presented in Table 2.[3] It was found that day school blind children scored significantly higher than both seeing and residential school blind subjects on the two subtests and on total score of the Word Fluency test. This is consistent with earlier findings of greater fluency in day school blind children (Kenmore, 1965; Lax, 1953; Maxfield, 1936; Payne, 1931). No differences in word fluency existed between the residential blind subjects and seeing subjects.

The seeing subjects scored significantly higher than the residential blind on all subtests of the Product Improvement test and significantly higher than the day school blind on the flexibility and originality dimensions. These differences could be a function of the test item itself. The test required that the student describe ways to improve the toy dog in order to make it more fun to play with. It is possible that seeing subjects were utilizing visual cues which triggered further responses, thus inflating their scores.

Of the 30 comparisons made on the Unusual Uses tests, 4 yielded significant differences, all of which were in favor of the seeing group. These subjects scored significantly higher than the day and residential blind subjects on both the flexibility and breadth dimensions of the Brick Subtest. The breadth score on this subtest reflected the number of different categories of uses

---

[3] Complete data on results of the statistical analyses are contained in the original report of the study (Tisdall, Blackhurst, & Marks, 1967).

TABLE 2
DIVERGENT THINKING TEST SCORE MEANS AND STANDARD DEVIATIONS

Test	Day school blind children		Residential school blind children		Sighted children	
	M	*SD*	*M*	*SD*	*M*	*SD*
Word Fluency						
B	17.29	6.28	14.97	6.37	14.53	4.20
T	18.53	7.18	16.13	6.51	16.47	4.27
Total	35.83	12.55	31.11	12.17	31.00	7.48
Product Improvement						
Fluency	10.61	6.18	9.19	5.49	11.96	6.26
Flexibility	4.44	3.06	4.05	2.73	5.69	2.97
Originality	16.04	10.35	14.72	9.54	19.97	11.50
Elaboration	12.71	8.08	10.61	6.33	13.56	7.43
Unusual Uses						
Brick						
Fluency	6.51	3.17	6.55	3.11	7.13	2.97
Flexibility	3.87	2.69	3.99	2.29	4.87	2.22
Breadth	3.73	1.72	3.76	1.48	4.59	1.42
Heat						
Fluency	8.25	3.76	7.53	3.30	8.31	3.02
Flexibility	5.35	2.76	5.05	2.56	5.67	2.24
Breadth	5.17	1.96	5.03	1.83	5.36	2.00
Water						
Fluency	7.12	3.68	7.35	3.24	6.92	2.54
Flexibility	4.81	2.96	5.11	2.62	4.88	2.10
Breadth	4.73	2.26	5.04	1.93	5.31	1.72
Total Fluency	21.00	8.85	21.44	7.99	22.43	7.35
Consequences						
Food						
Fluency	8.93	4.42	8.19	3.17	7.36	2.81
Flexibility	4.33	2.77	4.05	2.33	4.37	2.01
Breadth	3.68	1.50	3.43	1.38	3.75	1.14
Remote	1.60	1.72	1.56	1.62	1.97	1.83
Obvious	7.33	4.25	6.56	3.14	5.39	2.99
Read and write						
Fluency	8.85	3.44	8.47	3.65	7.69	3.02
Flexibility	3.95	2.59	3.51	2.67	3.60	2.03
Breadth	3.39	1.48	3.16	1.56	3.21	1.30
Remote	1.95	1.59	2.07	2.26	1.75	1.45
Obvious	6.89	3.06	5.40	2.94	6.01	3.00
Balance						
Fluency	6.13	3.29	6.79	3.42	6.45	2.68
Flexibility	3.48	2.29	3.61	2.35	3.89	2.28
Breadth	3.52	1.60	3.53	1.45	3.72	1.55
Remote	1.37	1.61	1.93	2.03	1.56	1.53
Obvious	4.76	2.95	4.87	2.65	4.89	2.19
Total Fluency	23.92	8.87	23.44	8.60	21.51	7.40
Ideational Fluency						
Drink	10.72	5.27	9.73	4.56	8.71	3.29
Sweet	8.72	5.07	7.79	5.22	6.37	3.78
Smooth	10.20	5.75	11.05	5.70	10.67	4.85
Green	5.79	5.37	6.00	4.26	4.17	3.61
Total	35.44	17.29	34.56	15.09	29.96	11.79
Seeing Problems						
Hammer	2.08	1.59	2.71	1.30	2.87	1.45
Wind	2.93	1.46	2.92	1.20	3.51	1.00
Glue	2.51	1.56	2.99	1.27	3.23	1.31

named, and flexibility was determined by the number of shifts between different categories. It could be postulated that seeing subjects scored higher on these items because they have seen bricks used in different ways in their environment. This might be construed as an accumulation of incidental knowledge that would not be available to the blind by virtue of their visual handicaps. However, it is interesting to note that the blind did not differ from the seeing children on the fluency score of the Brick item. This is fairly consistent with the earlier fluency findings and indicates that the blind were inferior in the quality, but not quantity, of their responses.

Sixteen scores were derived from the Consequences test battery. The day school blind children scored significantly higher than the seeing children on the fluency dimensions of the Food and Read-Writing Subtests and the obvious responses on the Food Subtest. Residential school blind children scored higher than the day school blind group on the remote score of the Balance Subtest, and higher from seeing children on the obvious score of the Food Subtest. All other differences in Consequences were nonsignificant.

Day school blind subjects scored significantly higher than seeing subjects on the Drink, Sweet, and Green Subtests, and on the total Ideational Fluency score. Residential blind subjects scored significantly higher than the seeing subjects on the Green Subtest and on the total ideational Fluency score. Thus, the greater fluency of the blind subjects was again in evidence.

On the Consequences test battery, seeing children scored significantly higher than both groups of blind children on the Wind Subtest. Both the seeing and residential school blind children obtained significantly higher scores on the Hammer and Glue subtests than did the day school blind subjects.

Seeing males scored significantly higher than females on 11 of the 41 tests of divergent thinking. Likewise, day school blind males scored significantly higher than females on eight of the subtests. No pattern for these differences emerged. The only instance of demonstrated superiority of females on the tests of divergent thinking was in the residential school group on the breadth dimension of the Unusual Uses-Brick subtest. All other differences in the residential group were nonsignificant.

An attempt was made to examine differences which may have existed in the divergent thinking abilities of children according to varying ages of onset of blindness. The range of onset of blindness could not be anticipated prior to the selection of the sample. Consequently, there was no initial guarantee that this variable would be manifest in enough subjects of various onset ages to extract meaningful information.

It was found that of the 152 blind subjects, 134 were congenitally blind, 10 were blinded between birth and 1 year of age, 3 became blind between the ages of one and five, and 5 became blind after the age of five. In attempting to analyze these data, it was decided that an insufficient number of adventitiously blind subjects precluded the drawing of sound conclusions regarding the divergent thinking abilities of the blind subjects. This decision was further supported by the fact that many of the children in the birth to 1 year onset range became blind within a few weeks or months of birth. The high incidence of congenitally blind children is explained by the etiological data which indicated that most of these subjects suffered retrolental fibroplasia.

On the basis of this investigation, it was concluded that blind children exhibit more verbal fluency in the divergent thinking dimension of intelligence than do seeing children. Visual familiarity with the environment, as opposed to inherent intellectual disparity, allows seeing children some advantage over blind children on a few divergent thinking tasks; however, more often than not, blind and seeing children in the 10 to 12 year age range do not differ in the ability to think divergently. Blind children in residential and day school settings are equally capable of thinking divergently; although there is a slight tendency for those in day school programs to be more verbally fluent.

When sex differences appear on the dimensions of divergent thinking, seeing and day school blind males tend to be more divergent than their female classmates. There are essentially no differences in the divergent thinking abilities of blind males and females who are enrolled in residential school programs.

REFERENCES

CHRISTENSEN, P. R., & GUILFORD, J. P. *Ideational fluency*. Beverly Hills, Calif.: Sheridan Supply Company, 1957.

CHRISTENSEN, P. R., & GUILFORD, J. P. *Word fluency*. Beverly Hills, Calif.: Sheridan Supply Company, 1958.

CHRISTENSEN, P. R., MERRIFIELD, P. R., & GUILFORD, J. P. *Consequences*. Beverly Hills, Calif.: Sheridan Supply Company, 1958.

GUILFORD, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1956.

GUILFORD, J. P. Traits of creativity. In H. H. Anderson (Ed.), *Creativity and its cultivation*. New York: Harper, 1959.

GUILFORD, J. P. *The structure of intellect model: Its uses and implications*. (University of Southern California Report from the Psychological Laboratory) April, 1960.

JONES, J. W. Problems in defining and classifying blindness. *Research Bulletin*. New York: American Foundation for the Blind, August, 1963, 123-129.

KENMORE, J. R. Associative learning by blind versus sighted children with words and objects differing in meaningfulness and identifiability without vision. Unpublished doctoral dissertation, University of Minnesota, 1965.

LAX, B. A. A comparison of the imaginative productions of congenitally blind and seeing children to structure auditory stimulation. Unpublished master's thesis, University of Texas, 1953.

LOWENFELD, V. *Creative and mental growth*. New York: MacMillan, 1952.

LOWENFELD, V. What is creative teaching? In E. P. Torrance (Ed.), *Creativity: Proceedings of the second Minnesota conference on gifted children*. Minneapolis: University of Minnesota Press, 1959.

LOWENFELD, V. Psychological problems of children with impaired vision. In W. M. Cruickshank (Ed.), *Psychology of exceptional children and youth*. Englewood Cliffs, N.J.: Prentice Hall, 1963.

MAXFIELD, K. E. The spoken language of the blind pre-school child, *Archives of Psychology*. 1936, **29**, 53.

PAYNE, S. Free association in blind children. Unpublished master's thesis, Columbia University, 1931.

SEEING PROBLEMS TEST, Aptitudes Research Project, Department of Psychology, University of Southern California, Los Angeles, California, Undated.

TISDALL, W. J., BLACKHURST, A. E., & MARKS, C. H. *Divergent Thinking in Blind Children*. United States Office of Education, Grant Number 32-27-0350-6003, 1967.

TORRANCE, E. P. & PALAMUTLU, N. Product improvement task. In E. P. Torrance, K. Yamamoto, D. Schenetzki, N. Palamutlu, and B. Luther (Eds.), *Assessing the creative thinking* *abilities of children.* Minneapolis: Bureau of Educational Research, University of Minnesota, 1960.

TORRANCE, E. P., YAMAMOTO, K., SCHENETZKI, D., PALAMUTLU, N., & LUTHER, B. *Assessing the creative thinking abilities of children.* Minneapolis: Bureau of Educational Research, University of Minnesota, 1960.

UNUSUAL USES TEST. Aptitudes Research Project, Department of Psychology, University of Southern California, Los Angeles, California, Undated.

WILSON, R. C., GUILFORD, J. P., & CHRISTENSEN, P. R. The Measurement of Individual Differences in Originality, *Psychology Bulletin*, 1953, **50**, 362–370.

(Received June 23, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 474–481

RESPONSE LATENCY:
ERRORS AND SUBSEQUENT CHANGE OF RESPONSE ON RETEST[1]

PHYLLIS W. BERMAN[2] AND MARY ANN VILLWOCK
Pennsylvania State University

The relationships between response latency, and (*a*) errors, and (*b*) change of response on retest, were investigated in a computer study with 53 college subjects. Two types of problems were presented in a program which was administered to each subject twice without feedback, and with a 1-week interval between tests. With both types of problems, (*a*) students made more errors and had longer latencies in response to problems which involved more, rather than less, response competition; and (*b*) the response latency for errors was longer than the latency for correct responses, even when significant problem variaables were weighted equally. Results were equivocal with respect to a third hypothesis: that latencies would be longer for responses which were changed on retest than for responses which were unchanged.

Recent advances in computer technology have made response latency a readily accessible measure of problem difficulty. At this time, however, far less is known about latency than about error rate, even though several writers (e.g., Brooks, Clark, & Park, 1967; Silberman, Melarango, Coulson, & Estavan, 1961; Suppes, 1964) have suggested that latency may provide us with information beyond that which is provided by error rate.

Although there has been an abundance of research relating latency to independent variables with simple laboratory learning tasks (Smith, 1968), there has been considerably less research exploring the possible use of latency with educational materials. A study by Brooks (1961) offers some exciting possibilities. Brooks measured the latencies of a small group of students responding to programmed instruction in remedial English. He found that longer latencies were associated with errors. Moreover, the average latencies of a small group of students on test items were related to the error rates of a large group given the same program. On this basis, he suggested that response latency might be used as a guide for the revision of programmed materials. The question which arises is to what extent we may generalize from Brooks' findings. Although many researchers have found longer latencies to be associated with item difficulty, high error rates, or more precriterion responses (Bluhm & Kennedy, 1965; Elias & White, 1969; Erickson, Zajkowski, & Ehmann, 1966; Fitts & Beiderman, 1965; Millward, 1964; Siegel, 1964), there have been some experimental situations in which shorter response latency has been associated with errors (Pollio & Reinhart, 1967; Rabbitt, 1966).

The present computer study was designed to investigate the relationship between response latency and the tendency to make an error, rather than a correct response, and the tendency to make a stable, rather than an unstable response. Problem difficulty was varied by designing material so that the amount of competition between responses would be varied. Two widely different types of materials were used. It was assumed that the difference in strength between competing response tendencies would be negatively related to the latency of the response and to the probability of response change on another occasion. That is, if there were little difference between two competing response tendencies, latency would be long, and the response would be likely to change from one occasion to another; if either of the response tendencies were much stronger than the other, response latency would be short, and the same response would be likely to recur on another occasion.

---

[1] This study was supported by United States Office of Education Contract OE-5-85-074, "Experimentation with Computer-Assisted Instruction in Technical Education." The authors are grateful to Karl Borman and Rosemary Hollick for their help in programming and analyzing the study, and to Keith Hall and Harold Mitzel for putting the Computer Assisted Instruction Laboratory facilities at their disposal.

[2] Requests for reprints should be sent to Phyllis W. Berman, Department of Psychology, 417 Psychology Building, Pennsylvania State University, University Park, Pennsylvania 16802.

It was assumed that the relationship between latency and errors would depend on the strength of the error response compared with that of the correct response. In most teaching situations the correct response tendency is usually stronger than, or almost as strong as, the error tendency. When there is little competition between responses the correct response is probably dominant. Therefore, a competitive situation with two strong response tendencies is likely to be associated with longer latencies and a higher probability of error than a situation in which there is little response competition. Less frequently, there are situations in which the incorrect response is dominant, and the converse would then be true.

The following hypotheses were proposed:

1. Students would make more errors and have longer response latencies on problems which involved a high degree of response competition than on those with little response competition.

2. Students would have longer test and retest response latencies for problems on which the response was changed from the test to the retest than for problems on which the response remained the same.

3. In the present study, problems were designed so that they would be of varying difficulty with both the error response tendency and the correct response tendency varying in strength. However, some instructions and examples of correct responses were given so that the correct response tendency was probably either stronger than the error tendency or, if it was weaker than the error tendency, it was at least moderately strong. It was therefore expected that the response latency for errors would be longer than the response latency for correct responses.

METHOD

Subjects

Fifty-three college undergraduates from two introductory educational psychology courses were enrolled in the experiment as paid volunteers. There were 30 men and 23 women. Students were told that they would be paid a flat rate for serving in the experiment in two sessions.

Procedures

The materials were presented to students by means of an IBM 1130-based 1500 instructional computer system. Responses were made via the keyboard or the light pen, constructed from a light-sensitive photocell. The same computer program was administered to students twice, with a 6 to 8-day interval between the two tests. Students were given instruction and practice using the light pen and typing and entering responses. They were told to answer as soon as they thought they knew the correct answer and that both speed and accuracy were important. The two keys used to enter responses were covered with red tape and referred to as the "red keys" in an attempt to facilitate discrimination and shorten response time. Introductory and instructional frames remained on the interface until the student indicated his readiness to proceed by pressing the red keys.

Materials

Two widely different types of problems were used. The material was chosen to be suitable for college undergraduates and to be somewhat independent of past training or coursework. It was desirable that students make errors on both the test and the retest. Therefore, some of the material was designed to be difficult, and only brief instruction and no feedback were given.

One of the problem types involved counting in bases other than Base 10. After receiving instructions and an example of a Base 3 series, the students were asked to furnish the next number in a Base 4, Base 5, or Base 6 series. In each of the bases, three types of questions were asked, and in each type of question an increasing amount of interference was expected from overlearned habits associated with Base 10.

In the first type of question, Type A, the correct response was the same as it would have been in the Base 10 system: Base 4: 1, 2, 3, 10, ——. Base 5: 1, 2, 3, 4, 10, 11, 12, 13, ——. Base 6: 1, 2, 3, 4, 5, 10, 11, 12, ——. In Type B questions the correct response differed from Base 10 in that the right digit changed to 0, and the left digit increased by 1, (e.g., 13 followed by 20 in the base 4 system): Base 4: 1, 2, 3, 10, 11, 12, 13, ——. Base 5: 1, 2, 3, 4, 10, 11, . . ., 33, 34, ——. Base 6: 1, 2, 3, 4, 5, 10, 11, . . ., 14, 15, ——. In Type C questions the correct response differed from Base 10 in that an extra place was added (33 followed by 100 in Base 4): Base 4: 1, 2, 3, 10, . . ., 30, 31, 32, 33, ——. Base 5: 1, 2, 3, 4, 10, 11, . . ., 40, 41, 42, 43, 44, ——. Base 6: 1, 2, 3, 4, 5, 10, 11, . . ., 50, 51, 52, 53, 54, 55, ——.

Ten set problems were also presented. In each problem, three geometric designs (set members), which had a critical feature in common, and three nonset members, without this critical feature, were shown. Students were asked to select from two alternatives the design which was a member of the set. The degree of interference from competing response tendencies was varied by varying the number of ways in which the alternative choice figures

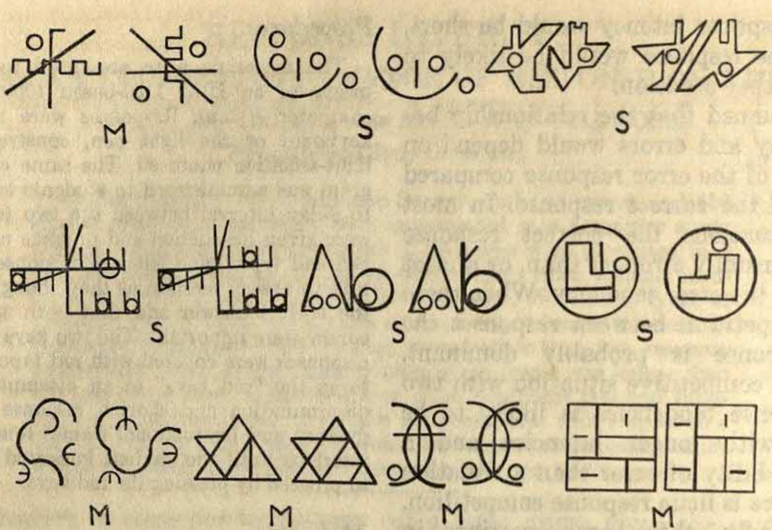


FIG. 1. The two choice figures for each of the five single-dimension (S) problems, and the five multidimension (M) problems.

differed from each other. Thus, in single-dimension problems alternative choice figures differed in only one aspect, the critical aspect; in multidimension problems the figures differed in the critical aspect and irrelevant aspects as well.

Ten groups of figures were designed. Two groups were modifications of figures in Foley (1968). The eight remaining groups were designed to be of medium to high complexity, with each figure composed of several contours. For each problem three examples of set members and three examples of nonset members were constructed. Five of the groups of figures were then randomly chosen for the single-dimension problems, and the remaining groups were assigned to the multidimension problems. The two alternative choice figures for each problem were constructed accordingly.

Each of the set problems was presented in two frames. In the first frame, the two choice figures were shown for 5 seconds. Students were asked to notice the difference between the figures. In the next frame, eight figures were shown: three set members, three nonset members, and the two choice figures which had been shown in the previous frame. The students were asked to point with the light pen to the choice figure which was a set member. Figure 1 shows the choice figures for each of the 10 set problems.

Before the 10 problems were presented, a simple practice problem was given to teach the method of comparing the two choice alternatives for salient features.

Orders

The problems were given in 24 different orders, all combinations of the following: set or base problems first, two orders of single-dimension (S), and multidimension (M) problems (M, S, S, M, M, S, S, M, M, S), and its inversion, and six orders of base problems (Bases 4, 5, 6; 5, 6, 4; 6, 4, 5; 4, 6, 5; 5, 4, 6; 6, 5, 4). In each of the bases the three types of questions were presented in a different order (Base 4: A, B, C; Base 5: B, C, A; Base 6: C, A, B). Two students each were given 1 of the 24 orders, and the additional five students were each given 1 of the orders.

RESULTS

There was some loss of data due to technical difficulties in recording some of the students responses and latencies with the result that only 47 students had complete data for the set problems, and only 28 students had complete data for the base problems. In each statistical test, data were included from all students with sufficient data for the particular analysis.

Comparison Between Base and Set Data

The combined test and retest error rates were remarkably alike for the base and set problems (22.3% for base problems and 20.5% for set problems), but students took more than three times as long to respond to set problems, and the difference was highly significant ($F = 108.868$, $df = 1/26$, $p < .001$). Both error rate and latency were significantly reduced on retest (error: $F = 9.494$, $df = 1/26$, $p < .005$; latency: $F = 22.615$, $df = 1/26$, $p < .001$). There were significant interactions between the problems and test–retest change in both the error rate and latency data, (error: $F = 6.117$, $df = 1/26$, $p < .05$; latency: $F = 10.249$, $df = 1/26$, $p < .005$). Between test and retest there was a large reduction in the error rate for base problems, but only a small reduction in error rate for set problems. On the other hand, there was a small reduction in response latency for base problems, but a large reduction in latency for set problems. Subjects' sex, order, (base or set problems first), and interactions with these factors were all insignificant sources of variance.

Relationship Between Response Measures for Students

Product-moment correlations between mean latency and mean error scores for individual subjects were calculated, using scores of the 47 students with complete set data and the 28 students with complete base data. Material with only moderate retest stability had been designed so that hypotheses concerning the latency of changed responses could be tested. Nevertheless, test–retest reliability for latency (bases: $r = .65$, $df = 26$, $p < .01$; sets: $r = 50$, $df = 45$, $p < .01$) compared favorably with reliability for error rate (bases: $r = .64$, $df = 26$, $p < .01$; sets: $r = .35$, $df = 45$, $p < .05$).

On the base problems, latency and error rate were not significantly correlated for test performance, but they were positively correlated for retest ($r = .54$, $df = 26$, $p < .01$). The opposite was true for the set data. That is, test latency and error rate were negatively correlated ($r = -.41$, $df = 45$, $p < .01$), but retest latency and error were not significantly related. Both response measures seemed to measure fairly specific abilities tapped by the two tests, and the correlations between error rate for base and set problems and between average latency for the two problems were not significant on either the test or retest.

The test latency and error data for both types of problems were also correlated with the only suitable measure of time to read and study instructions which had been included in the program: this was the reading time for the general instructions to the base problems. The frame included instructions for counting in Base 3 and showed a Base 3 series as an example. The frame remained on the interface until the student signaled that he was ready to proceed.

The base-instruction reading time was not significantly associated with base latency or error rate, but it was positively associated with set latency ($r = .55$, $df = 26$, $p < .01$) and negatively correlated with set error rate ($r = -.47$, $df = 26$, $p < .05$).

Effects of Independent Variables

The effects of the independent variables were tested with analyses of variance. Analyses of responses to the base problems were calculated separately for test and retest latency data and for error data, with test and retest combined.

The order in which the three base-system problems were administered had no significant effect upon the number of correct responses or response latency and therefore data for all students were combined.

For base problems, the particular base system used had no effect on students' scores or latencies, but the type of question asked (A, B, or C type) was a highly significant variable (number correct: $F = 13.360$, $df = 2/56$, $p < .001$; test latency: $F = 26.546$, $df = 2/52$, $p < .001$; retest latency: $F = 11.496$, $df = 2/52$, $p < .001$), with the most errors and the longest latencies for Type C questions, and the least errors and shortest latencies for Type A questions. Thus, the degree of competition from overlearned habits associated with Base 10 was a significant variable, but the particular base system used was not.

Analyses of the set data showed that more single-dimension than multidimension problems were solved correctly on both the test ($F = 48.615$, $df = 1/46$, $p < .001$) and the retest ($F = 63.222$, $df = 1/46$, $p < .001$). Although response latency was longer for multidimension problems on the test ($F = 7.98$, $df = 1/46$, $p < .01$), on the retest the difference in latency between single- and multidimension problems failed to reach significance.

There was also significant variation in the number of correct responses and latency among the five single-dimension problems and among the five multidimension problems (Berman & Villwock, 1969).

The two orders in which the set problems were given did not significantly affect response latency or interact with any other factor. Sex was not a significant variable for either type of problem. On base problems, sex interacted significantly with several variables for number correct or latency, but not both, and for test or retest, but not both (Berman & Villwock, 1969). However, no clear pattern emerged.

In summary, response competition did significantly affect the number of correct responses on the combined test and retest for the base problems and on both the test and the retest for the set problems. Latency was also significantly longer on both the test and the retest for the base problems when there was response competition. This was true for the set problems on the test, but not the retest. However, there was significant variation in latency and scores among both the multidimension problems and the single-dimension problems on the test and the retest. The base problem error rate and latency measures showed no significant variability due to the particular base system used.

Relationship Between Latency and Errors for Problems

The hypothesis that students would have longer response latencies for problems on which they made errors than for those on which they gave correct responses was tested by a t test. The data for both types of problems are presented in Table 1. Only students who had both correct and incorrect responses were included in each analysis. In calculating each student's average latencies, significant variables were weighted equally that is, test and retest, type of question for base problems, and multi- and single-dimension problems for set data.

TABLE 1

MEANS OF SUBJECTS' WEIGHTED AVERAGE LATENCIES (IN SECONDS) FOR CORRECT RESPONSES AND FOR ERRORS

Problem	Correct	Error	t	n
Base	9.7	12.9	3.06*	31
Set	31.4	47.1	4.90*	47

* $p = <.005$.

It took significantly longer for students to respond to problems on which they made errors than to problems which they answered correctly. This was true of both types of material. Inspection of the error and latency data for each of the 10 set problems revealed no obvious association between the two measures. On the other hand, there was a close relationship between error rate and average response latency for each of the nine base problems ($r = .82$, $df = 26$, $p < .005$; test data).

Errors on the base problems were sorted into several categories, and mean test latency for each type of error was calculated (counts as in Base 10, $M = 12.5$ seconds, $SD = 7.9$ seconds; fails to add extra place, $M = 13.0$ seconds, $SD = 5.3$ seconds; premature change, $M = 14.3$ seconds, $SD = 8.6$ seconds, unclassifiable errors, $M = 14.1$ seconds, $SD = 6.3$ seconds). The mean latency for correct test responses was 10.3 seconds, with a standard deviation of 3.9 seconds. The data are, of course, somewhat related to the significant differences between A, B, and C type questions since counting as in Base 10 would not be an error in Type A questions, and failing to add an extra place would not be an error in Type A and B questions. It is interesting, though, that unclassifiable errors, which appear on the surface to be accidents due to misreading of questions or carelessness, take as long or longer than other types of errors and only occur after the student has spent considerable time.

Latency for Changing and Unchanging Responses

The hypothesis concerning latency of changing versus unchanging responses was tested with a series of t tests. Base and set problems and test and retest problems were considered separately. Because incorrect responses had longer latencies than correct responses, correct responses were always compared with other correct responses, and errors were compared with errors. Thus, correct test responses which were later changed

TABLE 2

MEANS OF STUDENTS' WEIGHTED AVERAGE LATENCIES FOR RESPONSES TO PROBLEMS WHICH CHANGED FROM TEST TO RETEST AND THOSE WHICH DID NOT CHANGE

Problems	Unchanged	Changed	t	n	p
Set					
Latency of test responses					
Correct responses (correct to correct versus correct to error)	35.7	47.8	1.83	28	<.08
Incorrect responses (error to error versus error to correct)	34.0	39.2	.77	24	<.45
Latency of retest responses					
Correct responses (correct to correct versus error to correct)	22.6	32.6	2.74	34	<.025
Incorrect responses (error to error versus correct to error)	24.7	40.9	1.88	17	<.10
Base					
Latency of test responses					
Correct responses (correct to correct versus correct to error)	7.58	8.51	.63	7	<.60
Latency of retest responses					
Correct responses (correct to correct versus error to correct)	8.14	8.89	.42	15	<.70

to incorrect responses on retest were compared with those correct test responses which would remain correct on retest. Incorrect responses which would be changed were compared with those which would remain incorrect. Similarly, correct retest responses to problems which had been answered correctly on the test were compared with those which had been changed from incorrect answers, etc. Only students with both changing and unchanging responses were included in each analysis, and significant variables were weighted equally to determine each student's averages (i.e., single- and multidimension set problems and type of base question).

There were, then, eight categories in which the latencies of responses which changed could be tested against the latencies of those which did not change. However, there was a smaller number of students with base responses than set responses in each category, and only two students with sufficient data to compare incorrect test responses which changed on retest with those which did not change; similarly, there were only three students with sufficient data to compare incorrect base retest responses which had changed with those which had not changed.

Table 2 presents the means of students' weighted average latencies for changing and unchanging responses which belong in the six remaining categories. In all six of the response categories the differences were in the predicted direction; average latencies for responses which changed were longer than average latencies for responses which did not change. The probability of differences in this direction in six out of six possible categories is .016. However, within each category, differences were not very large. In three of the four categories of responses to set problems, the differences reached a probability level of less than .10, but in only one of these categories did the difference exceed the .05 level.

DISCUSSION

The results confirmed two of the study's hypotheses and were equivocal with respect to the third.

Students made more errors and had longer response latencies on problems which involved more competition between responses than on those with less response competition. This was true with both types of problems even though in each, response competition was varied in a different manner.

In the set tasks the number of responses which appeared relevant to problem solution were varied. Single-dimension problems could be solved by simply attending to the one feature of the stimuli on which the two choice figures differed. In contrast, the critical feature of the multidimension problems could be any of a number of dimensions on which the choice figures differed. In order to solve the multidimension problems students were required to attend to each of these stimulus features until they found the critical feature. In the base problems, the degree to which overlearned habits from Base 10 were in conflict with new responses associated with other base systems varied.

The results are perhaps more clear-cut for the base problems than for the set problems, since the difference between the response latencies of single and multidimensional set problems failed to reach significance on the retest. In constructing the set figures there were no controls for perceptual variables. Instead, perceptual differences were randomly distributed among single and multidimension problems since the 10 sets of stimuli were randomly assigned to the two conditions. Perceptual factors were undoubtedly a source of significant variability in errors and latency among the five single-dimension problems and among the five multidimension problems. A possible means of controlling for these differences would be to construct two sets of choice figures for each problem, thus creating a single and a multidimension form of each problem, each to be administered to a different group of students.

Students took longer to produce errors than correct responses on both types of problems. The differences were highly significant and were not due to the fact that errors and long latencies were both associated with significant factors such as test versus retest, single- versus multidimension problems, or type of base question asked. These factors had been equally weighted in calculating students' average latencies for correct and incorrect responses.

In the set problems two response alternatives were given; in the base problems, students produced a response from an unlimited number of alternatives, although the two strongest response tendencies were presumed to be the correct response and the response which would be correct in the Base 10 system. Of course, other responses were possible and were, in fact, made. It is interesting that these other error responses were also associated with longer latencies than correct responses.

Although base instructions were held to a minimum, they were obviously sufficient to establish the correct response as a strong competitive response. If it were not so, there would be little competition between the correct response and Base 10 habits, and "quick errors" based on strong Base 10 habits would probably be made. Pollio and Rinehart (1967) found such low latency errors. Their subjects learned to count in Bases 2, 3, and 4, without any preliminary instruction. After subjects produced each number in a series they were told the correct response. Subjects made many errors at the power changes (Type C problems). However, the time taken to produce numbers at power changes increased with each power change up to the third or fourth in the series. This increase in response latency coincided with an increase in the number of correct responses produced by the group.

The procedure used in the present experiment was more like the usual school situation in which some instructions are given, so that the correct response becomes a strong, if not dominant, response. However, it is well to remember that the positive association between error rate and latency may be restricted to situations where the error response is not in a strong dominant position.

In the set task there was no obvious relationship between error rate and mean latency for the 10 problems. The correlation between error rate and mean latency for the nine base problems was .82. Fitts and Beiderman (1965) found a correlation of .87 in a totally different type of task. Brooks (1961) with still another type of material, found correlations between the mean latencies of small groups and the error rate of a large group of students, which ranged between .58 and .83.

It should be noted that, while students took longer to produce error responses than correct responses, students who had longer mean response latencies did not necessarily make more errors than those with shorter mean latencies. There was a significant association on the base problems between students' mean latencies and error scores on the retest, but not the test. In contrast, the relationship between mean latency and error scores on the set problems was negative (and significant on the test, but not the retest). The relationship may indicate that more care was taken by students who were more successful, and this interpretation is supported by the fact that students who spent more time looking at the instructions for the *base* problems took longer to respond to *set* problems and made fewer errors on set problems. These results may be related to Kagan's distinction bwtween reflective and impulsive individuals and to his findings that impulsive children had faster response times and higher error scores on a battery of tests including several perceptual matching problems (Kagan, Pearson, & Welch, 1966).

It was expected that students would have longer test and retest response latencies for problems on which the response was changed from test to retest than for problems on which the response remained unchanged. Mean latencies of changed and unchanged responses within six categories were compared. In all six categories the average latencies of changed responses were longer than those of unchanged responses, but the difference between changed and unchanged responses reached significance in only one category. The number of students with appropriate data to make these tests was small, particularly for the base problems. This hypothesis cannot be accepted, based on the results of the present experiment, but it is certainly worthy of further testing.

REFERENCES

BERMAN, P. W., & VILLWOCK, M. A. Response latency: Its relationship with errors and response stability in two tasks involving response competition. In H. E. Mitzel (Ed.), Experimentation with computer assisted instruction in technical education. (Semi-annual Progress Report, No. R27, Project No. 5-85-074, United States Office of Education) University Park, Pa: Pennsylvania State University 1969.

BROOKS, L. O. Response latency in programmed learning: Latency related to error rate. Unpublished doctoral dissertation, University of Houston, 1961.

BROOKS, L. O., CLARK, J. W., & PARK, B. Response time as a function of response mode and item difficulty. (*AIR-C28-4/67-TR11.*) Palo Alto, Calif.: American Institute for Research, 1967.

BLUHM, P. M., & KENNEDY, W. A. Discrimination reaction time as a function of incentive-related DRQ anxiety and task difficulty. *Perceptual and Motor Skills,* 1965, **20,** 131–134.

ELIAS, M. F., & WHITE, S. H. Some psychophysiological and latency correlates of rule induction. *Psychonomic Science,* 1969, **16,** 84–85.

ERICKSON, J. R., ZAJKOWSKI, M., & EHMANN, E. D. All-or-none assumptions in concept identification: Analysis of latency data. *Journal of Experimental Psychology,* 1966, **72,** 690–697.

FITTS, P. M., & BEIDERMAN, I. S-R compatability and information reduction. *Journal of Experimental Psychology,* 1965, **69,** 408–412.

FOLEY, J. L. Sets, sub-sets, and operations. (Project Mathematics, ESEA Title III) West Palm Beach, Fla: 1968.

KAGAN, J., PEARSON, L., & WELCH, L. Conceptual impulsivity and inductive reasoning. *Child Development,* 1966, **37,** 583–594.

MILLWARD, R. Latency in a modified paired-associated learning experiment. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3,** 309–316.

POLLIO, H. R., & RINEHART, D. The acquisition of counting rules for some Non-Base 10 number systems. Paper presented at the meeting of the Psychonomic Society, Chicago, October 1967.

RABBITT, P. M. Errors and error correction in choice-response tasks. *Journal of Experimental Psychology.* 1966, **71,** 264–272.

SIEGEL, P. S. Concept formation: Response time considerations. *Psychological Reports,* 1964, **14,** 435–442.

SILBERMAN, H. F., MELARANGO, R. J., COULSON, J. E., & ESTAVAN, D. Fixed sequence versus branching autoinstructional methods. *Journal of Educational Psychology,* 1961, **52,** 166–72.

SMITH, E. M. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin,* 1968, **69,** 77–110.

SUPPES, P. Modern learning theory and the elementary school curriculum. *American Educational Research Journal,* 1964, **1,** 79–93.

(Received June 29, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 482–486

AUDITORY-VISUAL INTEGRATION, SEX, AND READING ACHIEVEMENT[1]

DAVID H. REILLY[2]
University of North Carolina at Chapel Hill

Research attempts to relate either auditory or visual discrimination skills to reading achievement have resulted in equivocal results. The present study was directed towards investigating auditory-visual integration skills as they related to sex and reading success. An N of 225 students in Grades 1–4 comprised the sample. Significant correlations were found at each grade level. The results suggested that females develop auditory-visual integration skills earlier than males. With females these skills appeared to reach an asymptote by the second grade. Sex differences in the developmental pattern of auditory-visual integration skills were suggested.

In the past few decades a substantial, if equivocal, body of research has been generated relating a child's ability to make auditory and visual discriminations to his reading achievement. Investigations have approached this issue from several viewpoints using various methodological strategies. The results appear to warrant the conclusion that there is a relationship between discrimination skill and reading achievement. The extent of this relationship and its specific nature remain unclear. One possible reason is that the majority of research in the area has concerned itself with discrimination skills rather than with the higher order integrative abilities which appear to influence learning success in a more specific manner. The current study was an effort towards investigating auditory-visual integration skills and reading achievement as they pertained to two selected variables, grade level and sex.

Initial attempts to relate auditory discrimination to reading skill began to appear in the 1920–1930 era. Among others, Deputy (1930), Monroe (1932), Bond (1935), Schonel (1948) and Templin (1943) have investigated these factors. A characteristic of the earlier studies seemed to be the expectation that a highly significant and direct relationship would be found between the two but this did not materialize.

Conclusions concerning visual discrimination and reading achievement are similar to those relating auditory discrimination to reading success (Muehl, 1960, 1961).

The combination of various approaches to auditory discrimination and reading skills have resulted in equivocal results and uncertainty as to the nature of their relationship. Thus far, predictive studies have been unable to appreciably affect the variance already established by mental age which itself can account for only less than half of the variance.

Katz (1967) concluded that better discrimination performance was associated with older children and with better readers at each grade level. This suggested that discrimination ability was a function of developmental factors.

Hirst (1969), in examining sex as a predictor of reading success, commented on the amount of research data indicating that girls were superior to boys in reading achievement. Birch and Belmont (1964) also noted this usual result when boys and girls were compared with regard to reading achievement.

The results of these studies suggest that sex differences, in addition to developmental stages, need to be considered when examining reading achievement.

More recent studies, concerned with auditory and visual discrimination and integration skills, have tended to focus on intersensory systems and reading success. This appears to be a more profitable approach since a basic lack in the area of auditory and visual discrimination and integration skills has been research combining the effects of both modalities and relating them to relevant areas of learning.

---

[1] This study was supported by Grant VA031-01-xxxx-2900 of the University Research Council of the University of North Carolina at Chapel Hill.

[2] Requests for reprints should be sent to David H. Reilly, School of Education, University of North Carolina, Chapel Hill, North Carolina 27514.

Birch and Belmont (1964) compared auditory-visual integration in good and poor readers. Their results indicated that good readers received higher auditory-visual integration scores than poor readers. Only boys born during a particular year were included in their study.

The present study hypothesized that auditory-visual integration ability was significantly related to reading achievement. Further, it hypothesized that, in investigating this relationship, consideration needed to be given to developmental stages and sex.

METHOD

Subjects

A total *N* of 225 comprised the sample of the current study. This included a random selection, using a table of random numbers, of eight classrooms from Grades 1–4 from a large city school system. The *N* of children for each grade level, by sex, is indicated in the Results section.

Procedure

All classrooms were administered the grade-appropriate form of the Gates-MacGinitie Reading Test. These forms were for Grade 1, Primary A; Grade 2, Primary B; Grade 3, Primary C; and for Grade 4, Survey D.

All classes were administered the reading test in groups. The testing was done by graduate students enrolled in a school psychology program.

To obtain an auditory-visual integration score (AVIS) the following procedure was utilized. All students in the sample were administered the AVIS test individually by one of the examiners. The test was an adaptation of the one used by Birch and Belmont (1964) in their study of auditory-visual integration and reading success. For the present study their procedure was modified as follows: The number of items was increased from 10 to 20 and the sequential difficulty of the items was altered appropriately.

The test calls for the examiner to strike a series of taps on a desk with a pencil, pen, or similar instrument, according to a planned sequence such as (.) (..) (..). The child has to listen to the taps and then pick the appropriate sequence from a series of three presented visually. The directions used for the current testing were the same as those described by Birch and Belmont (1964).

TABLE 2

F RATIOS FOR MALES AND FEMALES FOR GRADE LEVEL COMPARISONS OF AVIS, VOCABULARY, COMPREHENSION, AND TOTAL READING SCORE

Groups compared	AVIS	Vocabulary	Comprehension	Total reading
Total males–females	.77	5.06*	4.56*	5.66*
Grade 1	.49	2.50	4.15*	3.46
Grade 2	.39	2.55	3.86	3.28
Grade 3	.97	.01	0.00	.01
Grade 4	.05	1.00	2.95	2.25

* $p < .05$.

RESULTS

Table 1 presents the mean scores and standard deviations for the AVIS, as well as the raw scores obtained for the vocabulary, comprehension, and total reading scores from the Gates-MacGinitie test.

Table 2 presents a summary of the *F* ratios obtained from the analysis by sex of

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR AVIS, VOCABULARY, COMPREHENSION, AND TOTAL READING SCORES BY SEX AND GRADE LEVEL

Group	*n*	AVIS		Vocabulary		Comprehension		Total	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Total males	103	13.85	4.90	26.18	11.70	22.20	11.55	48.50	21.77
Total females	122	14.43	4.83	29.43	9.98	25.30	10.21	54.74	17.52
Males 1	31	8.90	3.05	24.74	13.27	15.03	8.51	39.81	21.04
Females 1	29	9.52	3.70	29.62	10.32	18.93	5.99	48.55	14.53
Males 2	20	12.65	3.67	24.00	12.38	16.75	9.26	40.75	21.00
Females 2	36	13.39	4.52	29.39	11.94	21.56	8.49	50.94	19.72
Males 3	23	16.17	4.03	32.78	10.46	27.83	10.53	61.04	20.95
Females 3	28	17.14	2.99	32.61	7.39	27.85	8.77	60.46	15.27
Males 4	29	18.13	2.03	24.00	8.59	29.17	10.40	53.21	18.14
Females 4	29	18.00	2.52	26.24	8.52	33.86	10.37	60.10	16.84

TABLE 3

F RATIOS FOR TOTAL BETWEEN-GRADE AND SAME SEX BETWEEN-GRADE ANALYSES

Grades	Total between grade			Same sex between grade					
				Males			Females		
	MS_b	MS_w	F	MS_b	MS_w	F	MS_b	MS_w	F
First–Second	446.23	14.45	33.87*	170.66	10.92	15.62*	240.76	17.46	13.79*
Second–Third	342.26	15.15	22.59*	132.84	14.97	8.87*	221.95	15.45	14.36*
Third–Fourth	50.45	8.45	5.96**	49.48	9.46	5.23**	10.47	7.63	1.37

* $p < .01$.
** $p < .05$.

the total population and for each grade level for the AVIS, vocabulary, comprehension, and total reading scores.

These results indicated that there was not a statistically significant difference between males and females on AVIS for the total sample or at any grade level tested. There was a statistically significant difference (.05 level) between total males and females with the vocabulary comprehension and total reading scores. A statistically significant difference was found between grade 1 males and females with the comprehension score. No other comparison by sex and grade level was significant at the .05 level. This result did not confirm the results of other investigators who have generally reported females as significantly higher than males in reading achievement. Reference to Table 1 indicates, however, that with the exception of the vocabulary and total reading scores at the third grade, the females did achieve consistently higher, though not statistically significant, scores.

TABLE 4

CORRELATIONS OF AVIS WITH VOCABULARY, COMPREHENSION, AND TOTAL READING BY GRADE AND SEX

Group	*n*	AVIS–Vocabulary	AVIS–Comprehension	AVIS–Total Reading
Total	225	.30*	.61**	.50**
Total males	103	.30*	.66**	.52**
Total females	122	.29*	.56**	.49*
Total Grade 1	60	.23*	.14	.21
Males	31	.12	.19	.15
Females	29	.34*	.04	.25
Total Grade 2	56	.65***	.71***	.70***
Males	20	.75***	.81***	.80***
Females	36	.61***	.68***	.67***
Total Grade 3	51	.42***	.34**	.40***
Males	23	.44*	.48**	.48**
Females	28	.40*	.18	.30
Total Grade 4	58	.25*	.30**	.30**
Males	29	.46*	.47**	.47**
Females	29	.11	.20	.17

* $p < .05$.
** $p < .01$.
*** $p < .005$.

Table 3 indicates the *F* ratios obtained when between-total grade level and same sex between-grade level analyses were computed.

The results obtained indicate significant differences between grade levels in auditory-visual integration skills. They also reflect significant differences between grade levels by sex, with the exception of the female groups at the third- and fourth-grade levels.

The findings are in keeping with Katz's (1967) conclusion that better discrimination performance is associated with older children. Although the present study was concerned with integration skills, it appears reasonable to assume that discrimination ability is an important, perhaps prerequisite, component of integration ability.

The results also suggest a modification of Birch and Belmont's (1965) conclusion that auditory-visual integration of temporal and spatial patterns reach an asymptote by the fifth grade. The current results suggest that sex differences need to be considered before their conclusion is accepted. It appears from the current results that females develop their auditory-visual integration skills to a higher level earlier than males, and that these skills reach an asymptote around the third or fourth grade. The current results do not allow a conclusion to be drawn about males, but the data suggest a trend towards reaching an asymptote at or above the fourth-grade level.

Table 4 presents the correlations of AVIS with vocabulary, comprehension, and total reading scores for total males and females, total grade level, and by sex for each grade level.

These results suggest that AVIS is a correlated function of reading achievement with variations in the degree of relationship dependent in part on sex and grade level.

For Grade 1, AVIS correlated significantly with vocabulary but not with comprehension or total reading score. For Grades 2–4, AVIS correlated significantly at the .05 level or better with the three areas of reading achievement examined.

Several trends related to sex and grade level were suggested. In the first grade no significant correlation was obtained between AVIS and the three areas of reading achievement for males. For the females, AVIS and vocabulary were significantly correlated.

For the second-grade data, significant correlations at the .005 level were obtained between AVIS and the three reading areas for both males and females.

The results from the third-grade data indicated a significant correlation between AVIS and the areas of reading for the males. For the female data, a significant correlation was obtained only between AVIS and vocabulary score.

For the males the fourth-grade data resulted in a significant correlation between AVIS and the three areas of reading. For the females, significant correlations were not obtained.

In general, the results suggested that the relationship between auditory-visual integration and reading achievement is related to grade level and sex. The relationship for both sexes appeared to be more pronounced in the second grade. For the males, the relationship appeared to be significant in Grades 2–4. For the females, it appeared to be significant in Grades 1–3.

DISCUSSION

In general, the results confirm previous findings that the ability to integrate auditory-visual stimuli is significantly related to reading achievement. This conclusion must be interpreted with regard for grade level and sex. The results suggest that females develop this ability to a higher level of success earlier than males, but that the ability reaches an asymptote earlier—probably by the fourth grade. For males, this ability appears to develop later, not reaching significance (Table 4) until the second grade and still evidenced at the fourth grade. For both males and females, the ability appeared to be most significant at the second grade. At this level it accounted for 56% to 64% of the variance with the females.

The conclusion of Birch and Belmont (1965) that auditory-visual integration of temporal and spatial patterns occurs between 5 and 7 years of age reaching an asymptote by the fifth grade may need to be modified to consider sex differences in the developmental pattern.

With comparisons made of the total male and female groups, auditory-visual integration ability appeared to be more closely related to comprehension scores than vocabulary accounting for approximately three times as much of the variance with comprehension as with vocabulary. Although word knowledge may be one prerequisite of comprehension the current results suggest that auditory-visual integration ability be given serious attention in any attempt to understand the means by which comprehension ability is developed. The data further suggest that teaching styles and remedial measures be further developed to consider this relationship.

The finding that auditory-visual integration skill development is related to both grade level and sex suggests that the ability needs to be carefully considered in developing appropriate and differentiated teaching techniques by grade level. It further suggests that consideration be given to developing appropriate and differentiated teaching techniques for males and females at the different grade levels with a particular focus on developing the auditory-visual integration skills of males at the kindergarten and first-grade level. The results suggest that the difference in reading achievement of males and females may be a function of auditory-visual integration skills.

REFERENCES

BIRCH, H. G., & BELMONT, L. Auditory-visual integration in normal and retarded readers. *American Journal of Orthopsychiatry,* 1964, **34,** 852–861.

BIRCH, H. G., & BELMONT, I. Auditory-visual integration, intelligence, and reading ability in school children. *Perceptual and Motor Skills,* 1965, **20,** 295–305.

BOND, G. The auditory and speech characteristics of poor readers. *Teachers College Contribution to Education,* No. 657, New York: Columbia University Press, 1935.

DEPUTY, E. C. Predicting first grade reading achievement: A study in reading readiness. *Teachers College Contributions to Education,* No. 426, New York: Columbia University Press, 1930.

HIRST, W. Sex as a predictor variable for success in first grade reading achievement. *Journal of Learning Disabilities,* 1969, **2,** 316–324.

KATZ, P. Verbal discrimination performance of disadvantaged children: Stimulus and subject variables. *Child Development,* 1967, **38,** 233–242.

MONROE, M. *Children who cannot read.* Chicago: University of Chicago Press, 1932.

MUEHL, S. The effects of visual discrimination pretraining on learning to read a vocabulary list in kindergarten children. *Journal of Educational Psychology,* 1960, **51,** 217–221.

MUEHL, S. The effects of visual discrimination pretraining with word and letter stimuli on learning to read a word list in kindergarten children. *Journal of Educational Psychology,* 1961, **52,** 215–221.

SCHONEL, F. J. *Backwardness in the basic subjects.* Edinburgh: Oliver and Boyd, 1948.

TEMPLIN, M. A study of sound discrimination ability of elementary school pupils. *Journal of Speech and Learning Disorders,* 1943, **8,** 127–132.

(Received July 31, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 487–491

STUDENT IMITATION OF A REWARDING TEACHER'S VERBAL STYLE AS A FUNCTION OF SEX AND GRADE LEVEL

PHILIP FRIEDMAN[1] AND NORMAN D. BOWERS
Northwestern University

Patterns of teacher and student verbal behavior within a framework of reinforcement theory were analyzed through the use of a modified form of the *Observation Schedule and Record 4V* (OScAR). The study examined the extent to which the rewarding teacher's classroom discourse served as a model for the verbal style of a student among his peers. Data were initially collected from 30 groups: 10 preschool, 10 kindergarten, and 10 first-grade classrooms. From each of these levels, six classes showing the highest frequency of teacher reinforcing verbal behavior were selected and designated as the "rewarding teacher" group. Verbal statements of four pupils (two boys, two girls) from each of these classes were scored using five scales of imitative behavior developed from the OScAR protocols. The data were analyzed using a 2 × 3 factorial design (Sex × Grade Levels). Female pupils imitated significantly more than males on four of the five verbal categories. In three of the analyses grade level differences were located, and for only one analysis was there a significant interaction.

A number of studies have demonstrated that children often modify their behavior as a result of simple exposure to a model (e.g., Bandura, 1962; Maccoby, 1959). Most of this research on the imitative behavior of children has been accomplished in a carefully controlled laboratory atmosphere. It would be of interest to determine if a child's verbal behavior could be changed by placing him in an ecologically natural situation where he has ample opportunity to observe, and perhaps to imitate, a verbal model. The preschool and elementary school classroom provides an excellent natural setting to study this issue, with the teacher becoming the model for imitation. The purpose of this study was to determine the amounts and kinds of verbal behavior elicited by the teacher which is subsequently dispensed by early school age children to each other within the classroom.

The Observation Schedule and Record 4V (OScAR), with some modifications, was employed in obtaining normative information about patterns of teacher and student verbal behavior within a framework of reinforcement theory. Several authors have contributed to a hypothesis that rewarding models are imitated to a greater extent than nonrewarding models, even though the imitation act may not in itself be reinforced. Mowrer's (1950; 1960) theory of imitation is based upon the principle of secondary reinforcement from classical learning theory. A child will continue to imitate a model's behavior on occasions where direct reinforcement is not provided, because stimuli associated with a history of reinforcements have acquired secondary reinforcng properties. Experimental support of this hypothesis is provided in a study by Bandura and Huston (1961) who found that preschool subjects who had been given considerable social reward during 30 minutes of play imitated both verbal and motor responses displayed by the model to a much greater extent than subjects experiencing a similar period of nonrewarding interaction with the model. Social power (Maccoby, 1959) and dependency (Sears, 1957) have been identified as important variables affecting the amount of imitation, with an individual often being chosen as a model simply because he has the *power* to provide or withhold reward. An observer may also vicariously experience any reinforcement perceived to be attained by the model, resulting in an increased likelihood of imitation without direct reward of observer behavior (Bandura, Ross, & Ross, 1961).

Model characteristics, type of behavior

---

[1] Requests for reprints should be sent to Philip Friedman, Speech Clinic, 1831 Hinman Avenue, Northwestern University, Evanston, Illinois 60201.

being modeled, and individual observer differences have been investigated as factors in the imitation process. Children tended to imitate a like-sex model more frequently than an opposite-sex model in various experimental situations (e.g., Bandura, Ross, & Ross, 1963a, 1963b; Hartup, 1962; Maccoby & Wilson, 1957). Mischel and Grusec (1966) showed that imitation was enhanced by additional reinforcement from a model, but the effect depended on the type of behavior: aversive or neutral. Boys generally showed more imitative physical aggression than girls, the difference attributed to the clearly sex-linked nature of the behavior (Bandura et al., 1961). One study (Rosenblith, 1959) concluded that attentiveness of the model, when compared to attention withdrawal, markedly increased imitation, but only in preschool girls. Reliable differences in modeling behavior have been found at various age levels (e.g., Schwartz, 1959). A typical result was that 4 year old boys and 6 year old girls imitated significantly more often than 4 year old girls and 6 year old boys (McDavid, 1959).

The findings reviewed above may be summarized in that they (*a*) support a reinforcement theory of imitation, and (*b*) suggest that there are individual and environmental differences which modify the effect of a reinforcing model. Along these lines, the following experiment was based on the assumption that a reinforcing teacher-model is particularly conducive to a pupil's imitation of his behavior. This assumption was made even though it was recognized that any particular child in a classroom may receive, directly, only a very small amount of the teacher's total reward output. The expectation was that young children manifest a wide variety of imitative behavior, and that developmental stages would be apparent. Hence, it was predicted that the early school years contain a period of marked increases in the pupil's use of imitated verbal behavior in his social interaction with his classmates. It was also expected that some differences between boys and girls would occur with respect to the utilization of the teacher's verbal characteristics.

METHOD

Subjects

The children observed were enrolled either in an urban preschool, or at the kindergarten or first-grade level of an inner city public school. The pool from which subjects were drawn consisted of 10 classes with 10 separate teachers at each of the three grade levels. In addition to the teacher, four students within each class were arbitrarily selected for observation. The total sample included 60 boys and 60 girls. Although there was some overlapping in the children's ages among the three grade levels, there was a mean difference of almost 8 months between the preschool and kindergarten groups, and slightly more than 6 months between the kindergarten and first grade classes.

Procedure

Observations were carried out over a 4-week period near the end of the school year. The observer always spent some time sitting in the room, prior to collecting any data, to allow the teacher and students to become accustomed to his presence. The teachers were informed by the observer beforehand that an experimenter would be in the room, but they were not given any indication of the type of data to be recorded. The main reason for this preobservational "conference" was to determine whether the teacher intended to allow the students to form groups and to work on their own sometime during the day's lesson.

A single observer using the OScAR 4V rated the teacher during three 4-minute sessions. On the same day, during a subsequent period of interstudent activity without direct teacher interaction, similar accounts of the verbal interaction occurring between a student and his peers were recorded. The recording schedule was a virtual duplicate of the OScAR 4V with the individual student being rated in exactly the same manner as the teacher, and his peers playing the role of the students. A small group containing at least four students (two males and two females) was carefully observed. The four children were rated in random order for 1-minute periods. Four 1-minute time segments were recorded for each student. In this manner a detailed account was attained of the child's verbal behavior, and that of any child with whom he interacted.

Similarity between the student's and teacher's verbal characteristics was employed as an operational definition of imitation. Five categories of verbal behavior were tabulated from both the student and teacher schedules (Bowers & Vogel, 1967; Cronbach, 1958).

I) *Substantive Ratio*—ratio of (substantive statements + substantive interchanges + describing + informing + problem structuring) to total.

II) *Continuing-Initiating Ratio*—ratio of continuing statements to initiating statements.

III) *Indirect-Direct Ratio*—ratio of (supportive + approving + accepting statements) to (nonevaluating + neutrally rejecting + rebuking + criticizing + directing statements).

IV) *Ratio Summation*—The sum of categories I, II, and III.

V) *Assumed Dissimilarity*—

$$D = \sqrt{(X_i/T)^2 - (Y_i/t)^2}.$$

A general measure of imitation; X represents a teacher statement, Y represents a statement of the observed students, T the total teacher verbal behavior, t the total student verbal behavior, and i any individual category on the two OScAR schedules. The degree of imitative behavior is defined as a reflection of the absolute difference of the ratios (Categories I–IV) from the two observation schedules, and the size of the assumed dissimilarity (Category V).

Analyses

The initial step in selecting the classes to be used in this experiment was a measurement of the reinforcement frequencies occurring among the teachers. The observation records were screened for instances in which the teacher verbally dispensed "social reinforcement" during the lesson. Four categories of positive social reinforcers were tabulated: (*a*) Considering—a statement revealing sensitivity to pupil feelings, (*b*) Supporting—a praising or enthusiastic response indicating the correctness of a student's behavior, (*c*) Approving—a nonenthusiastic response indicating acceptance of the student's behavior, and (*d*) Nonsubstantive pupil initiate interchange (positive)—a statement indicating that the teacher accepts, approves, or supports a pupil's suggestions (Medley, Impellitteri, & Smith, 1969). The ratio of the sum of these four categories divided by the total verbal record of the teacher was assumed to be an index of the model's frequency of reinforcement.

At each grade level, the six teachers with the highest scores on this reinforcement ratio were determined, and 72 of their students used for subsequent analysis. A 2 × 3 factorial design with 12 children of each sex at each of the three grade levels was employed.

RESULTS

The series of two-way analyses of variance, with each of the five categories of verbal behavior becoming a dependent variable, revealed some grade and sex differences. The Newman-Keuls test (Winer, 1962) was employed for multiple comparisons among the grade and sex levels when the main effects proved significant. For ratio summation (IV), there was a significant grade difference ($F = 8.21$, $df = 2/66$, $p < .01$). Individual comparisons revealed that first grade students showed significantly more imitative behavior than either kindergarten ($q = 3.80$, $df = 2/66$, $p < .01$) or preschool students ($q = 6.25$, $df = 3/66$, $p < .01$), while the difference between preschool and kindergarten failed to reach significance ($q = 2.09$, $df = 2/66$).

TABLE 1

ANALYSES OF VARIANCE OF IMITATION MEASURES

Source	*df*	Imitation scores				
		Substantive ratio	Continuing-Initiating ratio	Indirect-Direct ratio	Ratio summation	Assumed dissimilarity
Grade	2	13.34**	<1.00	<1.00	8.21**	3.82*
Sex	1	6.13*	1.03	5.32*	4.86*	5.70*
G × S	2	2.21	<1.00	4.36*	<1.00	<1.00
Within	66					

* $p < .05$.
** $p < .01$.

This observed grade difference is somewhat misleading as evidenced by an analysis of the ratio summary into its more specific components of verbal behavior (see Table 1). Most of the difference is accounted for by Category I (substantive ratio) ($F = 13.34$, $df = 2/66$, $p < .01$), with neither of the other two ratios reaching a significant level (see Table 1). Using the substantive ratio, the increase in imitative behavior the higher the grade level is most evident ($q = 5.57$, $df = 2/66$, $p < .01$, comparing the kindergarten and first grade classes; $q = 5.04$, $df = 2.66$, $p < 0.1$, comparing the preschool and kindergarten classes). The second general measure of imitative behavior, assumed dissimilarity, also revealed a significant main effect of grade level ($F = 3.82$, $df = 2/66$, $p < .05$). With this measure, however, no statistical distinction could be made between the kindergarten and preschool groups ($q = 1.67$, $df = 2/66$).

There were significant sex differences showing greater verbal imitation by girls in four categories, with only the continuing–initiating ratio failing to reach a significant level (see Table 1). For Category III (indirect–direct ratio) a significant Grade × Sex interaction was found ($F = 4.36$, $df = 2/66$, $p < .05$), with the difference favoring imitation by female students increasing sharply with grade level.

DISCUSSION

The analysis presented above suggests an increase in a student's imitation of the verbal behavior of his teacher from the preschool to the first grade levels. Although we have considered the mean age differences of the children in the three grades, a more important variable may be the increase in social experience. Along these lines, the findings coincide with the conclusions of Charlesworth and Hartup (1967) and others concerning the positive correlation between the degree of social participation and age level. The present findings, however, reveal that some verbal characteristics of a teacher may be more susceptible to imitation than others.

The analysis revealed that imitation of the teacher's substantive ratio, which is a measure of the proportion of his statements referring directly to classroom content and containing no affect, is the greatest source of difference among the three grade levels. One may ask if this is really an index of imitative behavior. It may be argued that all we have demonstrated is that the first grade teacher tends to set up a more formal classroom atmosphere than the kindergarten or preschool teacher. In the first grade the teacher places a greater emphasis on problem structuring, course work, and the conveyance of factual material, and less emphasis on the game and role playing of the kindergarten. Consequently, the peer group conversation should be more substantive, and closer to the typical verbal characteristics of a teacher.

A number of factors point to the inadequacy of this argument as an interpretation of the data. A simple comparison of the average substantive ratio's of the students at the kindergarten and first grades showed no significant difference ($t = .84$, $df = 46$). Furthermore, the variance of this substantive ratio among classes *at the same grade level* is very large for both students and teachers. Finally, despite the fact that the mean substantive ratio for the six preschool teachers was significantly greater than that for the kindergarten teachers ($t = 2.09$, $p < .05$), our data indicate more similarity between the kindergarten teachers and their students on this verbal category ($q = 5.04$, $df = 2/66$, $p < .01$).

Our most impressive result is the evidence for an early formation of sex differences in imitating the teacher's verbal traits. Two additional factors must be considered in interpreting these results. First, all of the 18 teachers included in our final sample were women. Since the sex of the model has been demonstrated time and again to be significant in both the identification and imitation processes (e.g., see Sears, 1953; Mussen & Rutherford, 1963; Hartup, 1962), the teachers' femininity probably plays a great role in explaining the pupil differences. Hence, any general conclusion of sex differences in the utilization of imitative behavior will depend on additional research with male teachers. Second, we may question to what extent these sex differences in imitation are independent of differences in the general level of social activity between young boys and girls. Charlesworth and Hartup (1967), for example, found that boys participated in considerably more "give and take play" in the nursery school than girls. Along these lines, it is interesting to note that in terms of what may be called "total verbal activity" (total score on all the OScAR categories), the mean frequency for boys was slightly higher than for girls, although this was not significant ($t = .47$, $df = 70$). This would indicate some degree of independence between our measures of imitation and quantity of verbal interaction.

The extension of the OScAR 4V observational technique has yielded promising results in measuring student as well as teacher verbal behavior, and the factor of imitation as one form of interaction between the two. The results of the study imply individual differences in opportunity and ability to develop identifications with the teacher. While a student's contact with his teacher lacks the intensity of his contacts with his parents or peers, there is every reason to believe that the direct observation of student-teacher interactions can contribute to a broader conception of classroom learning.

REFERENCES

BANDURA, A. Social learning through imitation. *Nebraska Symposium on Motivation*, Lincoln, Nebraska: University of Nebraska Press, 1962.

BANDURA, A., & HUSTON, A. C. Identification as a process of incidental learning. *Journal of Abnormal and Social Psychology*, 1961, **63,** 311–318.

BANDURA, A., ROSS, D., & ROSS, S. A. Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology*, 1961, **63,** 575–582.

BANDURA, A., ROSS, D., & ROSS, S. A. A comparative test of the status envy, social power, and secondary reinforcement theories of identification learning. *Journal of Abnormal and Social Psychology*, 1963, **67,** 527–534. (a)

BANDURA, A., ROSS, D., & ROSS, S. A. Imitation of film-mediated aggressive models. *Journal of Abnormal and Social Psychology*, 1963, **66,** 3–11. (b)

BOWERS, N. D., & VOGEL, F. X. Scales of teacher behavior derived from two methods of classroom observation. Paper presented at annual meeting of the American Educational Research Association, Chicago, February, 1967.

CHARLESWORTH, R., & HARTUP, W. W. Positive social reinforcement in the nursery school peer group. *Child Development*, 1967, **38,** 992–1002.

CRONBACH, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person, perception, and interpersonal behavior*, Stanford: Stanford University Press, 1958.

HARTUP, W. W. Some correlates of parental imitation in young children. *Child Development*, 1962, **33,** 85–96.

MACCOBY, E. E. Role-taking in childhood and its consequences for social learning. *Child Development*, 1959, **30,** 239–252.

MACCOBY, E. E., & WILSON, W. C. Identification and observational learning from films. *Journal of Abnormal and Social Psychology*, 1957, **55,** 76–87.

MCDAVID, J. W. Imitative behavior in preschool children. *Psychological Monographs*, 1959, **73,** 486.

MEDLEY, D. M., IMPELLITTERI, J. T., & SMITH, L. H. *Coding Teachers Verbal Behavior in the Classroom: A Manual for Users of OScAR 4V*. New York: City University of New York, Division of Teacher Education, 1969. (mimeo.)

MISCHEL, W., & GRUSEC, J. Determinants of the rehearsal and transmission of neutral and aversive behaviors. *Journal of Personality and Social Psychology*, 1966, **3,** 197–206.

MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald Press, 1950.

MOWRER, O. H. *Learning theory and the symbolic processes*. New York: Wiley, 1960.

MUSSEN, P. H., & RUTHERFORD, E. Parent-child relations and parental personality in relation to young children's sex role preferences. *Child Development*, 1963, **34,** 589–607.

PARTEN, M. B. Social participation among preschool children. *Journal of Abnormal and Social Psychology*, 1932, **27,** 243–269.

ROSENBLITH, J. F. Learning by imitation in kindergarten children. *Child Development*, 1959, **30,** 69–80.

SCHWARTZ, N. An experimental study of imitation. The effects of rewards and age. In J. F. Rosenblith, Learning by imitation in kindergarten children. *Child Development*, 1959, **30,** 69–80.

SEARS, P. S. Child rearing factors related to the playing of sex-typed roles. *American Psychologist*, 1953, **8,** 431. (Abstract).

SEARS, R. R. Identification as a form of behavior development. In D. B. Harris (Ed.), *The concept of development*. Minneapolis: University of Minnesota Press, 1957.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw Hill, 1962.

(Received July 10, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 492-500

PERCEPTUAL SPEED IN POOR READERS[1]

CARL SPRING[2]
Stanford University

Poor and normal readers, matched on age (8-11), IQ (91-141), and sex were tested to determine if poor readers process visually presented letters slower than normal readers. Two upper-case letters were presented simultaneously. Subjects were required to respond by pressing one switch if the letters were the same, and by pressing another switch if the letters were different. Subjects were instructed to respond as fast as possible, while minimizing errors. It was found that latencies were longer for poor than normal readers. In addition, an estimate of central-processing time was obtained by subtracting latency to judge that two highly dissimilar letters were different from latency to judge that two slightly dissimilar letters were different. It was found that central-processing time was significantly greater for poor than normal readers, only for data collected in the second half of the test sessions. It appeared, therefore, that central-processing speed of poor readers deteriorated significantly during a period of approximately 10 minutes of intensive testing.

Children who cannot learn to read, or who learn to read only with great difficulty, may have normal intelligence, adequate cultural stimulation, and normal hearing and vision. They may present no gross neurological signs, and be free of apparent emotional stress. The source of the problem, therefore, is somewhat obscure.

Previous studies have attempted to identify reading disability with impairment of visual perception. Benton (1962), in a review of investigations of visual perception related to reading disability, concluded that if there is a relationship, it is weak, and appears to be noticeable only with young children. Typically, studies of the relationship between reading disability and visual perception have used error-frequency as a measure of perceptual performance. The research reported here, however, focuses on perceptual speed. The hypothesis tested was that children who are poor readers process visually presented letters slower than normal readers matched on age, IQ, and sex.

The following method was used. Two upper-case letters were presented simultaneously, and subjects were required to respond by pressing one switch if the letters were the same, and by pressing another switch if the letters were different. Subjects were instructed to respond as fast as possible, while minimizing errors. Reaction time (RT) and error data were recorded. Inasmuch as there are peripheral as well as central components of total RT, however, straightforward measurement of RT could be grossly misleading in testing an hypothesis about central processes. It was necessary, therefore, to use a method of estimating the central components of RT. The method that was used is based on the widely replicated finding that, if stimuli are very similar, the latency to perceive their difference is greater than if they are highly dissimilar (Woodworth & Schlosberg, 1954). Since it is likely that this difference is entirely accounted for by differences in the amount of central sensory and decision processing required under conditions of varying similarity, the difference is assumed to be an estimate of pure central-processing time. In the present study, therefore, central-processing time was estimated from trials on which subjects correctly responded that paired letters, of varying degrees of similarity, were different. The estimate was obtained by subtracting total time to judge that two highly dissimilar letters were different from total time to judge that two slightly dissimilar letters were different. This method assumes that RT may be divided into a series of components which are additive. The assumption of additivity dates back to the work of Donders (1868), who attempted to measure the duration of certain RT components by comparing mean RTs from two different tasks, where one task is thought to require all the components of the first, plus an additional component. Several recent investigators have studied RTs as sums of component latencies, and some evidence that this is not an unreasonable assumption has been compiled (Hohle, 1967; Sternberg, 1969).

---

[1] This paper is based on a dissertation submitted to Stanford University in partial fulfillment of the requirements for the PhD degree. The author wishes to thank Richard E. Snow, Richard C. Atkinson, and Leo Ganz for their guidance.

[2] Requests for reprints should be sent to Carl Spring, who is now at the Department of Education, College of Letters and Science, University of California, Davis, California 95616.

METHOD

Subjects

Two groups of children, matched on age (8–11), IQ (91–141), and sex, were selected. The children were enrolled in Grades 3 through 5 of the Menlo Park School District, serving a middle to upper-middle class community. One group ($N = 25$), designated dyslexic, was about 2.5 years behind the other ($N = 25$), designated normal, in reading performance. Grade level of reading ability was determined from an average of the tests in the reading section of the Stanford Achievement Test, excluding the comprehension test.

An attempt was made to screen out emotionally disturbed children, based on teachers' assessments. Two children, one from each group, were replaced for this reason. All of the poor readers were enrolled in a special program for educationally handicapped children, and as a consequence had been tested for auditory and visual acuity. None of these children were known to have impaired hearing or an uncorrected visual impairment.

Intelligence was determined from available test results. The task of matching the dyslexic and normal groups on IQ was complicated by the fact that scores from different tests were available for different children. Most of the dyslexic children had been given the Wechsler Intelligence Scale for Children (WISC); although a few had been given the Stanford-Binet Intelligence Scale (SB). Most of the normal children, on the other hand, had Lorge-Thorndike scores; although several had WISC or SB scores. After giving careful consideration to differences in test content and methods of standardization, it was concluded that the dyslexic and normal groups could be fairly matched on IQ using these available scores. It appears that the three tests overlap considerably in the abilities they measure. This conclusion is supported by reported correlations in the area of .80 between WISC full-scale scores and Stanford-Binet scores (Krugman, 1951), and in the area of .70 between Lorge-Thorndike scores and either WISC or Stanford-Binet scores (Freeman, 1959). In addition, methods of standardizing the WISC, Stanford-Binet, and Lorge-Thorndike tests are similar. Each has been standardized on a carefully selected population, and reports IQ as a deviation score with a mean of 100. The WISC uses a standard deviation of 15, the Lorge-Thorndike uses a standard deviation of 16, and the Stanford-Binet uses a standard deviation ranging from 12.5 to 20 depending on age. These differences were not considered serious because (*a*) the difference between WISC and Lorge-Thorndike standard deviations is too small to distort comparisons seriously, and (*b*) Stanford-Binet scores are used for only six out of 50 children in the study, and these six children are distributed equally between poor and normal readers.

Of additional interest is the fact that 72% of the dyslexic group, and consequently of the normal group also, were boys. This is consistent with other reports in the literature. There is agreement that the incidence of reading disability among males is at least twice that among females (Money, 1962).

Materials

As previously noted, the method for separating central-processing time from total RT was based on the widely replicated finding that the similarity of two stimuli, and the latency to judge that they are different, are correlated. It was, therefore, necessary to devise a metric by which the similarity of a pair of upper-case letters could be quantified. This method was based on a system of distinctive features by which the letters might be recognized. Examples of features from the feature list used in this study are: curvilinearity, oblique linearity, and vertical linearity. In all, the list included eight distinctive features. For each of these features, a value of zero or one was assigned to each upper-case letter used in the displays, representing the absence or presence of that feature in the letter. Thus, each letter was coded by an eight-bit binary code. With this code it is possible to express the similarity of a pair of letters as a function of the number of features on which one letter has a value of 0, while the other letter has a value of 1.

The feature list used in this study was derived from the writer's factor analysis of a matrix of perceptual-confusion frequencies, for upper-case letters, reported by Gibson, Osser, Schiff, and Smith (1963). On each trial they presented a target letter, followed by an array of several letters including the target; and preschool children were required to point to the target in the array. Gibson et al. reported two confusion-frequency matrices for the 26 upper-case letters, after folding by adding corresponding elements above and below the diagonal. These two matrices were collected under slightly different conditions, but were highly correlated. For the present study, the matrices were combined by adding. Resulting error frequencies were transformed, maintaining ratio intervals, to

TABLE 1

ROTATED FACTOR MATRIX AND COMMUNALITIES OBTAINED FOR THE FIRST EIGHT FACTORS, FROM LETTER-CONFUSION DATA ADOPTED FROM GIBSON, OSSER, SCHIFF, AND SMITH (1963)

Letters	Rotated factors								
	I	II	III	IV	V	VI	VII	VIII	h^2
A	.02	−.01	−.06	.03	−.02	.00	.04	.26	.08
B	.00	.16	−.01	.08	−.05	−.09	.37	−.01	.18
C	.00	.66	−.02	−.06	−.02	.10	.02	.02	.45
D	.03	−.06	.03	.07	.02	.50	−.02	−.04	.27
E	−.02	.03	−.04	.75	.01	−.02	.01	.00	.56
F	−.04	.07	.04	.66	.02	−.01	.02	−.05	.45
G	.00	.63	.06	−.05	−.04	.02	.23	−.11	.47
H	.11	−.05	.02	.39	−.02	.08	.06	.09	.19
I	.00	−.02	.01	−.03	.60	.01	.04	−.07	.37
J	.02	.38	−.05	.07	.05	.04	−.14	.09	.19
K	.05	.00	.70	.05	−.02	.03	.00	−.07	.50
L	−.02	.05	.10	.04	.68	−.02	−.03	−.06	.48
M	.74	−.01	−.02	.01	.00	.00	.02	.02	.55
N	.62	.06	.09	.06	.00	.01	−.03	−.03	.41
O	−.02	.06	−.01	−.03	−.01	.68	−.03	.03	.47
P	.01	−.11	−.05	.00	.08	.13	.62	.05	.42
Q	−.01	.05	−.01	.01	−.01	.65	.05	.01	.43
R	.01	.01	.00	.01	.01	−.03	.72	.04	.52
S	−.01	.30	.09	.04	.00	−.09	.03	−.01	.11
T	.02	−.01	−.05	.00	.55	.01	.01	.12	.32
U	.03	.27	−.20	.00	.07	−.01	−.14	.43	.32
V	.00	−.01	.07	.00	−.01	−.01	.00	.70	.50
W	.64	−.03	.00	−.01	.00	.00	.02	.02	.41
X	.02	−.01	.70	−.02	−.03	−.01	.00	.11	.50
Y	−.05	−.06	.28	.00	.01	.01	.06	.58	.43
Z	.00	.13	.23	−.01	.12	.01	−.05	.00	.09

bring each error frequency within the range: 0–1. This was done by dividing each frequency by 100. This transformed matrix, with 1s in the diagonal, was factor analyzed. The computer program (Dixon, 1967) computed a principal components solution with a Varimax rotation of the factor matrix. Four rotated factor matrices were obtained by rotating two, four, six, and eight factors. The rotated matrix, obtained by rotating eight factors, is shown in Table 1. Each rotated factor consists of a cluster of letters which tend to be confused with each other. Clusters of letters, obtained by rotating two, four, six, and eight factors, are illustrated in Figure 1.

Figure 1 reveals a hierarchical order among the clusters of letters. Violations of the hierarchical order are illustrated by dotted lines. The results of rotating two factors are illustrated at Level 1 of the hierarchy, four factors at Level 2, six factors at Level 3, and eight factors at Level 4. In general, as the analysis progresses from Level 1 to Level 4, large clusters split into smaller clusters. Similar results were obtained by Gibson, Schapiro, and Yonas (1968) in a study that used latencies to judge whether pairs of capital letters were same or different. A cluster analysis of their data revealed an underlying hierarchical structure similar to that illustrated in Figure 1.

From Figure 1 it was possible to derive a distinctive feature list with eight binary-valued features. Each feature is described by a property that is common to one of the letter clusters in Figure 1. The features are identified here:

DF1 (*H K X M N W*): The letters are concave up and down.

DF2 (*M N W*): The maximum number of points of intersection with any horizontal line is three or more.

DF3 (*A K V X Y Z*): The letters include oblique lines.

DF4 (*C D G J O Q S U B P R*): The letters include curved lines.

DF5 (*B P R E F H I L T*): The letters include vertical lines.

DF6 (*D O Q*): The letters include enclosed areas.

DF7 (*I L T*): The maximum number of points of intersection with any horizontal and any vertical line is one, disregarding serifs.

DF8 (*A U V Y*): The letters are concave up or down, but not both. Using this feature-list, an eight-bit binary code was assigned to each capital letter used in the present study. It should be

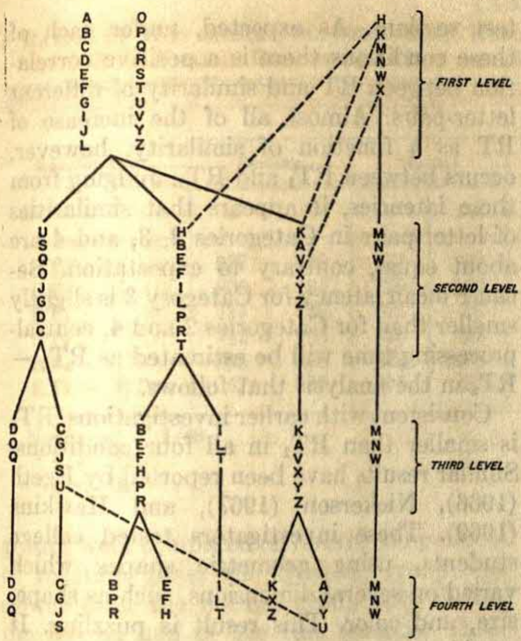


FIG. 1. Clusters of upper-case letters obtained from factor analysis of a confusion matrix adapted from a report by Gibson, Osser, Schiff, and Smith (1963).

noted, however, that very similar letters, *M* and *W* for instance, may have the same eight-bit code, because the feature list does not specify enough information to uniquely code every letter.

Letter pairs were typed in a sans-serif font, photographed as positive transparencies, and displayed by a slide-projector. Thirty-six "different" pairs, listed in Table 2, were used. These pairs of letters are divided into four categories, ranging from most to least similar according to the similarity metric described above. The most similar letter pairs are shown in Category 1, the next most similar in Category 2, etc. In addition, 36 "same" pairs were constructed from the 18 letters in the different pairs, using each letter in two same pairs. Thus, equal numbers of different and same pairs were used.

TABLE 2

THIRTY-SIX PAIRS OF UPPER-CASE LETTERS, RANGING FROM MOST SIMILAR (CATEGORY 1) TO LEAST SIMILAR (CATEGORY 4)

Category I	Category 2	Category 3	Category 4
O Q	C F	L U	U N
C G	F B	Y C	M O
M N	H T	P Y	P W
E F	O U	Y B	Q M
D P	F D	G Y	W D
B P	G E	T U	N O
I L	H I	O H	B W
T L	D E	Q H	U M
I T	E Y	W C	N Q

Procedure

Subjects were tested individually. The sequence of events was as follows. First, the subject signed his name, and it was noted whether signatures were made with the right or left hand. Two dyslexic and three normal children were left-handed. Next, instructions were read, followed by 28 practice trials and 72 test trials. Each trial required about six seconds, and a session lasted about 10 minutes, excluding instructions.

The 72 test trials consisted of 36 same and 36 different letter-pairs, randomly ordered. Order of presentation was reversed for half of the dyslexic and normal subjects. On each trial, two upper-case letters were presented simultaneously, and the subject was required to respond by pressing one of two hand-held microswitches, depending on whether the letters were the same or different. Subjects were instructed to respond as fast as possible, while minimizing errors. Displays were not terminated until the subject responded. Displays were preceded by a warning signal: the word "ready." In addition to this verbal signal, the sound of the Kodak Carousel projector, as it initiated the movement to advance to the next display, was a warning signal. This movement occupied 1 second, providing a fixed foreperiod.

Subjects were instructed to operate the same switch with their dominant hands. This procedure, of course, fails to protect against the possibility that same responses might be faster than different responses merely because RT is smaller when the dominant hand is used. This possibility was not considered serious for two reasons. In their review of the effect of motor factors on RT, Woodworth and Schlosberg (1954) stated that RT varies only within 5 milliseconds when different hands are used, and this difference does not consistently favor the dominant hand in all subjects. Furthermore, the possibility that the dominant hand might respond a few milliseconds faster is unrelated to the data that are critical to the hypothesis under investigation.

Displays were rear-projected onto a frosted screen in a room with normal lighting. The screen was masked by black cardboard with a 2 × 2½ inch rectangular display window. Luminescense of a display was roughly 30 foot Lamberts. The two letters of a pair were arrayed side-by-side. Subjects were seated about 6 feet from the display window. At this distance, a letter-pair subtended a visual angle of 5° from top to bottom edge, and 1.0° from left to right edge.

A light-activated switch, mounted behind the screen, started an electronic timer when a display was initiated. The timer was stopped when the subject pressed one of the two hand-held microswitches. In addition, each hand-held microswitch operated one of two colored lights mounted in a special unit. Response latencies to the nearest millisecond and errors were recorded.

RESULTS

Reaction Times

Before reporting results, it will be useful to introduce some notation. The symbols RT_s, RT_1, RT_2, RT_3, and RT_4 denote mean reaction times, in milliseconds, to give correct responses to stimulus displays in the corresponding Categories S, 1, 2, 3, and 4. Category S contains same letter-pairs; while Categories 1, 2, 3, and 4 contain different letter-pairs. According to the distinctive-feature list used in this study, these different pairs range from very similar (Category 1) to very dissimilar (Category 4).

Figure 2 depicts values of RT_s, RT_1, RT_2, RT_3, and RT_4 for the following four conditions: dyslexic and normal groups, subdivided into first and second halves of the test sessions. As expected, under each of these conditions there is a positive correlation between RT and similarity of different letter-pairs. Almost all of the increase of RT as a function of similarity, however, occurs between RT_1 and RT_2. Judging from these latencies, it appears that similarities of letter-pairs in Categories 2, 3, and 4 are about equal, contrary to expectation. Because mean latency for Category 3 is slightly smaller than for Categories 2 and 4, central-processing time will be estimated as $RT_1 - RT_3$ in the analysis that follows.

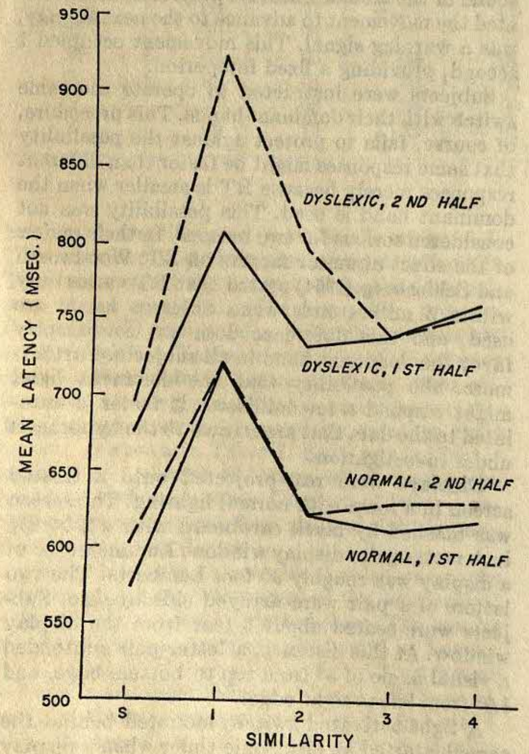


FIG. 2. Mean reaction times in milliseconds for five degrees of similarity of upper-case letter pairs. (Dyslexic and normal subjects, first and second halves.)

Consistent with earlier investigations, RT_s is smaller than RT_1 in all four conditions. Similar results have been reported by Egeth (1966), Nickerson (1967), and Hawkins (1969). These investigators tested college students, using geometric shapes which varied on several dimensions, such as shape, size, and color. This result is puzzling. It would seem that subjects must compare same pairs on all features or dimensions before initiating a same response; whereas comparisons of different pairs are apparently self-terminating when a difference is found. Thus, RT_s would be expected to be larger, not smaller, than RT_1. The present study sheds no light on this problem, except to observe that the result is obtained with letters as well as geometric figures, and with dyslexic as well as normal subjects.

In addition to these general observations that apply equally to both groups, it is apparent from Figure 2 that there are differences between the dyslexic and normal groups. First, total RTs are larger for the dyslexic group in both halves of the test sessions. Second, $RT_1 - RT_3$, the estimate of central-processing time, is larger for the dyslexic group, but only in the *second half* of the test sessions. Differences between dyslexic and normal groups on RT_s, RT_1, RT_2, RT_3, RT_4, and $RT_1 - RT_3$ were evaluated by *F* tests. Means, standard deviations, *F* ratios, and significance levels are shown in Table 3 for both halves of the testing sessions.

These data may be summarized as follows. When dyslexic and normal groups were compared on *total* reaction time, the dyslexic group was slower than the normal group. But when dyslexic and normal

TABLE 3

F RATIOS OF DIFFERENCES BETWEEN MEAN REACTION TIMES (IN MILLISECONDS) OF DYSLEXIC AND NORMAL GROUPS FOR FIRST AND SECOND HALVES OF TESTING SESSIONS

Reaction time	First half					Second half				
	Dyslexic		Normal		*F*	Dyslexic		Normal		*F*
	M	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*	
RT_s	677	123	573	69	13.5**	696	131	601	83	9.4**
RT_1	809	184	720	189	2.9	921	207	721	107	18.4**
RT_2	734	138	623	89	11.4**	801	239	623	95	11.9**
RT_3	742	158	616	97	11.5**	741	147	632	102	9.3**
RT_4	765	198	622	108	10.0**	759	176	635	103	9.1**
$RT_1 - RT_s$	67	147	103	132	.8	180	158	89	103	5.8*

Note.—*df* = 1/48.
* $p < .025$.
** $p < .01$.

groups were compared on *central* components of reaction time, the dyslexic group was slower than the normal group only during the second half of the test sessions. Thus, central components of latency, for the typical dyslexic child, deteriorated during 10 minutes of intensive testing; while central components of latency, for the typical normal child, were unchanged.

Errors

Egeth and Smith (1967) have proposed that errors in choice reaction-time tasks occur when subjects terminate processing and respond prematurely. This model predicts, in the present experiment, that Category 1 displays should be associated with the largest error frequencies. This is because a response is more likely to be premature, given that processing is arbitrarily terminated, if letter-pairs are very similar and require more time to process. Relative error frequencies, shown in Table 4, confirm this prediction. Relative error frequencies were calculated by dividing the number of errors in Categories S, 1, 2, 3, and 4 by the number of trials in each category.

Let us assume that premature responses result in correct, as well as erroneous choices. Premature error responses, called "detected errors" in the discussion that follows, were discarded before computing mean RTs. Premature correct responses, called "undetected errors," were not discarded, of course. It must be determined, therefore, if the results reported in the previous section are merely due to the possibility that the effect of undetected errors was greater for normal than for dyslexic subjects.

The net effect of undetected errors on mean RT depends on two factors. One factor is the probability of an undetected error, which cannot be directly estimated, and which is assumed in this analysis to equal the frequency of detected errors (Table 4). This is quite a strong assumption, for there is no compelling reason to believe that undetected errors occur as frequently as detected errors, or even that they occur at all. Another factor is the mean latency of undetected errors, which also cannot be directly measured, and which is assumed in this analysis to equal the mean latency of detected errors (Table 5).

It is instructive to compare mean latencies of detected errors, in Table 5, with mean latencies of correct responses (including undetected errors), in Table 3. In 17 out of

TABLE 4

RELATIVE ERROR FREQUENCIES

Stimulus category	First Half		Second half	
	Dyslexic	Normal	Dyslexic	Normal
S	.05	.08	.06	.07
1	.20	.28	.16	.19
2	.04	.03	.01	.05
3	.05	.09	.08	.04
4	.06	.06	.02	.05

TABLE 5

MEAN ERROR LATENCIES (IN MILLISECONDS)

Stimulus category	First half		Second half	
	Dyslexic	Normal	Dyslexic	Normal
S	556	501	747	597
1	565	495	654	539
2	696	473	912	580
3	636	600	572	570
4	569	406	955	630

20 possible comparisons, latencies of detected errors are smaller than latencies of correct responses (p = .001). This result is in accord with the error model proposed by Egeth and Smith, which attributes errors to premature termination of processing.

Given the data in Tables 3, 4, and 5, it is possible to estimate corrected mean latencies, uncontaminated by undetected errors. If the probability of an undetected error (p) is assumed equal to the relative frequency of detected errors, the probability of a correct response, excluding undetected errors, is $1 - 2p$. Given this assumption, it is possible to compute the weights of undetected-error responses and correct responses in mean latency. The weight of undetected errors is:

$$\frac{p}{(1 - 2p) + p}$$

which reduces to

$$\frac{p}{1 - p}.$$

The weight of correct responses is:

$$\frac{1 - 2p}{(1 - 2p) + p},$$

which reduces to

$$\frac{1 - 2p}{1 - p}.$$

Since uncorrected mean latency (L_u) is defined, in the error model discussed above, as the weighted sum of the mean latency of correct responses (L_c) and the mean latency of undetected error responses (L_e), the following formula may be derived:

$$L_u = L_c \frac{1 - 2p}{1 - p} + L_e \frac{p}{1 - p}.$$

Solving this equation for L_c, the following equation is obtained:

$$L_c = \frac{L_u - p(L_u + L_e)}{1 - 2p},$$

where L_c is the corrected mean latency; L_u is the uncorrected mean latency (Table 3); L_e is the mean latency of undetected errors, assumed equal to the mean latency of detected errors (Table 5); and p is the probability of an undetected error, assumed equal to the relative frequency of detected errors (Table 4).

Applying this formula, the corrected estimates of mean latency, shown in Table 6, were obtained. It may be seen that corrected RTs are larger for the dyslexic group than for the normal group in all 10 possible comparisons in Table 6 (p = .001). Thus, even after correcting for the contamination of undetected errors on the calculation of mean RT, it appears that dyslexic subjects were slower than normal subjects.

Central-processing time, estimated by $RT_1 - RT_3$, is now considered. For the dyslexic group, $RT_1 - RT_3$ was 142 milliseconds in the first half of the test sessions, and 227 milliseconds in the second half. These corrected data parallel those obtained before applying the correction formula, and are in accord with the conclusion that the performance of dyslexic subjects deteriorated during testing. The corrected data for normal subjects, however, do not parallel those obtained before applying the correction formula. Corrected central-processing

TABLE 6

CORRECTED MEAN REACTION TIMES (IN MILLISECONDS)

Stimulus category	First half		Second half	
	Dyslexic	Normal	Dyslexic	Normal
S	684	580	693	601
1	890	863	984	777
2	736	628	800	625
3	748	618	757	635
4	778	637	755	635

time improved, for normal subjects, from the first half to the second half of testing; and it was larger for normal than for dyslexic subjects in the first half, although the difference was reversed in the second half. One must remember, in evaluating these corrected estimates of central-processing time, that they are based on strong assumptions, and little error data.

DISCUSSION

That total RT was longer for dyslexic than normal children is of considerable interest. This large difference, on the order of 100 milliseconds, suggests the possibility of using a modification of the task used in this study for the early prediction of reading disability.

Of greater importance for understanding the nature of dyslexia, however, may be the present finding that the central-processing component of total RT increased significantly, during 10 minutes of intensive testing. This is similar to results obtained by Lynn (1960) and Otto and Fredricks (1963). In these studies, children were categorized into groups of good and poor readers, and tested with an inverted-digit printing task. In this task, children are required to print inverted numbers as rapidly as possible. These investigators found that performance of poor readers equaled that of good readers at the beginning of a test session. As the session continued, however, the performance of poor readers began to fall behind that of good readers. After a rest period of about 5 minutes, the performance of poor readers again equaled that of good readers.

It is tempting to think of the shift from normal to long perceptual latencies, seen in the present study, as a *cause* of reading disability. Such an interpretation, however, is not adequately supported by evidence. Until stronger evidence is available, the latency-shift described in this report must be regarded as a *correlate* of reading disability, along with several others. It is known, for example, that auditory perception is impaired in many poor readers; thus, impairment of auditory perception is another correlate of reading disability. The possibility that there is something rather fundamental about a shift from normal to long perceptual latencies, in certain cases of reading disorder, would be strengthened if this shift could be demonstrated for auditory as well as visual perception. This, and other questions about the relation of latency-shift to reading disability, will be taken up in future investigations.

REFERENCES

BENTON, A. L. Dyslexia in relation to form perception and directional sense. In J. Money (Ed.), *Reading disability*. Baltimore: Johns Hopkins Press, 1962.

DIXON, W. J., (Ed.). *Biomedical computer programs*. Los Angeles: University of California Press, 1967.

DONDERS, F. C. Die Schnelligkeit psychischer processes. *Archives of Anatomical Physiology*, 1868, 657–681.

EGETH, H. E. Parallel versus serial processes in multidimensional stimulus discrimination. *Perception and Psychophysics*, 1966, **1**, 245–252.

EGETH, H., & SMITH, E. On the nature of errors in a choice reaction task. *Psychonomic Science*, 1967, **8**, 345–346.

FREEMAN, F. S. The Lorge-Thorndike Tests. In O. K. Buros (Ed.). *The Fifth Mental Measurement Yearbook*. Highland Park, N. J.: Gryphon Press, 1959.

GIBSON, E. J., OSSER, H., SCHIFF, W., & SMITH, J. An analysis of critical features of letters, tested by a confusion matrix. (Final report on Cooperative Research Project No. 639, United States Office of Education, Department of Health, Education, and Welfare) Ithaca, N. Y.: Cornell University 1963.

GIBSON, E. J., SCHAPIRO, F., & YONAS, A. Confusion matrices for graphic patterns obtained with a latency measure. In, *The analysis of reading skill: A program of basic and applied research*. (Cornell University and United States Office of Education, Final Report, Project No. 5-1213) Ithaca, N. Y.: Cornell University, 1968.

HAWKINS, H. L. Parallel processing in complex visual discrimination. *Perception and Psychophysics*, 1969, **5**, 56–64.

HOHLE, R. H. Component process latencies in reaction times of children and adults. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior*, Vol. 3. New York: Academic Press, 1967.

KRUGMAN, J. L., & OTHERS. Pupil functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children. *Journal of Consulting Psychology*, 1951, **15**, 475–483.

LYNN, R. Individual differences in introversion-extraversion, reactive inhibition, and reading attainment. *Journal of Educational Psychology*, 1960, **51**, 318–321.

MONEY, J. Dyslexia: A postconference review. In

J. Money (Ed.), *Reading disability*. Baltimore: Johns Hopkins Press, 1962.

NICKERSON, R. S. "Same–different" response times with multi-attribute stimulus differences. *Perceptual and Motor Skills*, 1967, **24**, 543–554.

OTTO, W., & FREDRICKS, R. C. Relationship of reactive inhibition to reading skill attainment. *Journal of Educational Psychology*, 1963, **54**, 227–230.

STERNBERG, S. The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II. Acta Psychologica*, 1969, **30**, 276–315.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, Rinehart & Winston, 1954.

(Received July 13, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 501–505

VISUAL SEARCH AND READING[1]

ROBERT C. CALFEE[2] AND PENNY JAMESON
University of Wisconsin—Madison

Subjects were asked to look for occurrences of words from a target list while reading a passage for comprehension. Number of target words in memory (one, two, or four) was combined factorially with number of occurrences of target words in text (zero, one, two, or four) in a within-subjects design. In Experiment I ($n = 12$), subjects searched for the specific words in the target list, and in Experiment II ($n = 12$), for any close associate of words in the target list. In both experiments, reading speed was affected only by the number of words in the target list. Neither variable had any effect on comprehension. The results indicate that central memory load limits reading speed, but scanning and decoding processes are so automated that they are unaffected by tallying operations.

The tasks of visual scanning and memory search exemplified in experiments by Neisser (1967, Ch. 5) and Sternberg (1969) resemble in some ways the process of reading in literate adults. As an example of a study along this line, Neisser and Beller (1965) asked subjects to scan a list for a meaningful word (i.e., Monday) or for words falling in conceptual classes (such as states of the Union, proper names, or animals). They found that the rate of search for all words in a closed conceptual class was slower than for a single word—subjects searched through a list at a rate of .07 seconds/word when instructed to look for Monday, and .11 seconds/word when looking for the name of any of the states. Rate of search for animals (an open conceptual class) was even slower, about .18 seconds/word.

In reading, a subject can use context to predict particular words or classes of words which are likely to occur. To the extent that a skilled reader makes such predictions, rapid scanning for the "target" words should be possible, thereby facilitating speed of reading. Visual search studies typically have used lists of letters or single words as stimulus materials, and hence the results are not directly applicable to the processes involved in normal reading.

The present experiment was designed to investigate variables similar to those manipulated in visual search studies but in a task more like normal reading. The question of primary interest concerned the relation between reading speed, the number of target items for which the subject was searching, and the number of target occurrences in text.

METHOD

Subjects were presented a list of target words, and then read a passage which contained occurrences of target words (Experiment I) or close associates of the target words (Experiment II). There were 12 subjects in each experiment. They were instructed to notice these occurrences while reading the passage as rapidly as possible. Comprehension of the passage was stressed. As soon as the passage was finished, the page was turned and the reading time was recorded by the experimenter. The subject wrote down the frequency of occurrence of each of the target words and then answered two multiple-choice questions testing comprehension of the passage.

Design and Procedure

The basic design in both experiments was a 3 × 4 factorial within-subjects design. The variables were (*a*) number of target words in the memory list (one, two, or four), and (*b*) number of occurrences of targets in the passage (zero, one, two, or four). For conditions with one word in the memory list and two or four occurrences in text, the target word was repeated in the passage two or four times, respectively. For conditions with two words in the memory list and with two or four occurrences in text, each target word occurred once or twice, respectively. For the condition with four words in

---

[1] This research was supported in part by the Wisconsin Research and Development Center for Cognitive Learning with funds from the United States Office of Education Contract OE 5-10-154, and in part by Grant MH 12637 to the first author from the National Institute of Mental Health.

[2] Requests for reprints should be sent to Robert C. Calfee, who is now at the School of Education, Stanford University, Stanford, California 94305.

the memory list and four occurrences in the passage, each target word occurred in the passage exactly once. For conditions with fewer occurrences than memory words, no target occurred more than once in the passage; otherwise words were chosen at random from the target list.

The 12 conditions from the 3 × 4 factorial design were combined with 12 passages in a Graeco-Latin square, which was administered to two groups of 12 college undergraduates, volunteers from introductory psychology classes at the University of Wisconsin. Experiment I was conducted in the fall of 1967, and Experiment II in the Spring of 1968. In Experiment II, a no-target control list was incorporated in a partly counterbalanced fashion in the Graeco-Latin square.

The passages were chosen from familiar works of fiction, for example, *All the King's Men* by R. P. Warren, *Death in Venice* by Thomas Mann, *The Ballad of the Sad Cafe* by Carson McCullers, and *You Only Live Twice* by Ian Fleming. All passages were altered as necessary until they were approximately 100 words in length, not counting function words. Passages were also altered so that one common noun appeared four times in the course of the 100 words, and two common nouns twice each. Passages were originally selected to meet these requirements as closely as possible, and then nouns were substituted for pronouns where necessary. The passages were modified somewhat after Experiment I in an effort to render them more homogeneous, but as will be seen below these efforts were not completely successful.

The nouns that appeared as targets in Experiment I were high-frequency English words (Thorndike-Lorge AA or A). In conditions with more target words than occurrences, nonoccurring targets were compatible with the general sense of the passage, but no target or close associate of a target appeared in text. In Experiment II, the memory list shown to the subject before he read the passage consisted of close associates of the occurrences in the passage, and the subject was instructed accordingly. The occurrences in text were the same as the target words in Experiment I. The memory words were of lower frequency than the words occurring in text, but were still reasonably common (at least 50 times per million, Thorndike-Lorge).

The subjects were fully instructed as to the nature of the task. They were asked also for their grade point average, to report any special reading training courses, and to rate themselves both as to speed and accuracy on normal reading. None of these measures bore any significant relation to performance on the task, and they will not be discussed further. Subjects were then given practice runs on three lists similar to those used in the experiment proper. Following each practice list, they were asked if they had any questions. The 12 experimental passages were then given without any breaks. A session usually took less than half an hour.

Reading time was measured in centiseconds from the time subjects turned the page to begin reading until the page was turned over. On the test page which followed, the target list was reproduced, together with two multiple-choice questions. Subjects wrote down the number of times they thought each of the target words occurred, and then answered two questions. In Experiment II, the words listed on the test page were those actually occurring in text rather than the associates in the target list to insure that the subject knew exactly which words to tally.

RESULTS

Three dependent measures were analyzed: (*a*) reading time, (*b*) the difference between the number of words occurring and number of words reported, and (*c*) number of correct answers on the comprehension test. Table 1 contains the means for the first two dependent variables, and in Table 2 the analysis of variance of the data from Experiment I. The corresponding results for Experiment II are in Tables 3 and 4. In general, the two experiments yield nearly identical results. Reading speed is slower in Experiment II, and while it is tempting to speculate on the implications of this difference, it is true that the two studies were conducted in different semesters and hence possibly involved different populations. In any event, a plausible interpretation is that the additional burden of storing "associates" rather than explicit representations in memory is responsible for the slower reading rate. This interpretation is compatible with other aspects of the data.

TABLE 1

MEAN READING TIME (IN SECONDS) AND, IN PARENTHESES, NUMBER TARGETS REPORTED MINUS NUMBER TARGETS OCCURRING FOR EACH EXPERIMENTAL CONDITION IN EXPERIMENT I

Memory load	Target occurrences				
	0	1	2	4	Average
1	30.47 (.17)	29.50 (.17)	30.39 (−.50)	29.28 (−.25)	29.99 (−.10)
2	32.37 (.25)	31.61 (−.17)	32.70 (.17)	32.58 (−.27)	32.32 (.01)
4	34.05 (1.42)	33.10 (1.25)	33.24 (1.33)	33.86 (−.08)	33.56 (1.02)
Average	32.29 (.61)	31.40 (.42)	32.11 (.33)	31.90 (−.20)	31.96 (.31)

Note.—$N = 12$.

The most interesting question concerns the relation between reading speed, memory load, and number of target items in the text. In both experiments, reading time increased significantly ($p < .001$) with increased memory load (the number of different items subjects had to keep in mind while reading). The increase in reading time between one and four memory words was about 10% in both experiments. Reading time for the no-target control passages in Experiment II was 40.7 seconds, 8.1 seconds less than with one target in memory, which is further evidence of the effect of memory load on reading speed. However, performance was not affected by the number of occurrences of target words that the subject had to process or identify while reading, nor was there any evidence of an interaction between memory load and number of occurrences. The natural interpretation is that subjects were able to identify and tally target items without any disturbance of the reading process. Moreover, subjects performed these identifications with equal facility whether searching for specific target words or for close associates.

From the difference between the number of targets occurring and number of occurrences reported by the subject, shown in parentheses in Tables 1 and 3, it can be seen that memory load was also the primary variable in determining number of words reported. In general, subjects tended to report more words as the number of target words (and, hence, test words) increased. Considering just the condition with four words in the memory list but no occurrences in the text, six subjects in Experiment I and eight in Experiment II checked one or more of the test words as having been detected. When the memory list contained two words and there were no occurrences, two subjects in Experiment I and seven in Experiment II checked one of the test words, and when the target list contained one word and there were no target occurrences, one subject in Experiment I and two in Experiment II checked the test word. In short, the basic pattern of results for the number of words reported was this—subjects reported more words as memory load (and hence number of test words) increased, but this effect was negligible by the time there were as many as four occurrences of target items in text. This result is probably due to the fact that nonoccurring targets were selected to be compatible with the sense of the passage.

TABLE 2

ANALYSES OF VARIANCE ON DEPENDENT VARIABLES IN EXPERIMENT I

Source	*df*	Reading time	Number Seen—Number Occurring
		F	*F*
Subjects	11	27.7***	7.0***
Order	11	<1	2.0*
Passage	11	4.4**	1.1
Conditions			
Memory load (ML)	2	8.1***	16.8***
Number of occurrences (NO)	3	<1	4.5*
ML × NO	6	<1	1.4
Residual error (mean square)	99	14.21	1.03

* $p < .05$.
** $p < .01$.
*** $p < .001$.

TABLE 3

MEAN READING TIME (IN SECONDS) AND, IN PARENTHESES, NUMBER TARGETS REPORTED MINUS NUMBER TARGETS OCCURRING FOR EACH EXPERIMENTAL CONDITION IN EXPERIMENT II

Memory load	Target occurrences				
	0	1	2	4	Average
1	49.69 (.17)	50.19 (.17)	44.95 (−.25)	50.53 (−.58)	48.84 (.29)
2	52.73 (.58)	52.12 (.17)	54.77 (−.58)	53.06 (−.08)	53.17 (.35)
4	57.50 (1.00)	53.62 (1.33)	55.30 (.67)	53.33 (.00)	54.94 (.75)
Average	53.31 (.58)	51.98 (.56)	51.67 (.50)	52.31 (.22)	52.32 (.47)

Note.—$N = 12$.

Analysis of variance of the comprehension scores revealed no significant effects of experimental conditions, nor did visual examination of the means suggest any pattern to the comprehension data. Some of the passages were more difficult to comprehend than others, and some took longer to read than others. However, there was no correlation between comprehension scores and reading time for the different passages.

TABLE 4

ANALYSES OF VARIANCE ON DEPENDENT VARIABLES IN EXPERIMENT II

Source	*df*	Reading time	Number seen—Number occurring
		F	*F*
Subjects	11	15.0***	<1
Order	11	2.1*	1.5
Passage	11	6.7***	2.6*
Conditions			
Memory load (ML)	2	28.6***	13.7***
Number of occurrences (NO)	3	<1	3.4*
ML × NO	6	<1	2.4*
Residual error (mean square)	99	66.15	.87

* $p <$ 05.
** $p < .01$.
*** $p < .001$.

DISCUSSION

The finding that memory load was a significant factor in reading speed suggests that central, cognitive processes involving the storage or maintenance of readily available information plays a critical role in performance of skilled readers. The fact that the number of occurrences of target items in text was not related to performance indicates that the processes involved in perceptual scanning and preliminary decoding are so rapid, so automated, that recognition and tallying of words in text can be carried out without significantly impeding the reading act. Finally, the fact that these results hold whether a subject is looking for a word or some associate of a word suggests that the high-speed portion of skilled reading involves some amount of semantic decoding.

The data are consistent with a serial processing model of the following sort: as the subject reads a passage, chunks of information (probably in the form of strings of words which comprise coherent phrases) are read in at a constant rate. Each chunk of information is checked against the memory list, one word at a time in a serial, exhaustive fashion. Parenthetically, Sternberg (1969) has presented considerable evidence supporting the hypothesis of serial, exhaustive memory search in a variety of simpler experimental contexts. When there is a match between the information in a chunk and any of the target words, the subject makes note of this, but no additional time is required by the tallying operation. However, the rate at which chunks are read in does depend on the amount of time required to complete the memory search process.

While the model above provides a reasonable account of the data, the fact that the number of occurrences in text had no effect whatsoever on reading speed is still surprising, particularly in Experiment II. The rationale for instructing the subject to search for associates was that the tallying operation would require a substantial amount of time under these conditions—the memory search process, operating at high speed, might signal the subject that a tentative match had been detected, but then a more careful comparison would have to be made between the prospective associate in the text and the memory target. The additional time required to check out positive occurrences in text would thus lead to increased reading time as the number of occurrences in text became greater. However, this increase did not occur, nor was there much difference in memory search rate between the two experiments.

It is also difficult to reconcile the preceding interpretation with the negligible effects of memory load on comprehension scores. One might suppose that comprehension would require some of the same processing operations as memory storage, and hence would be influenced by memory load. One possibility is that the subject arranged a tradeoff—reading speed was adjusted as necessary to maintain a satisfactory level of comprehension. Another and more likely possibility is that the comprehension test procedure was not sufficiently sensitive, either because it was too easy or because the errors only reflected faulty test items. The passages were relatively short, and so there was not a great deal of content to measure. The multiple-choice tests used were aimed at specific facts in text, and the distractors selected to be reasonably similar to the correct answer. In Experiment II, subjects were asked to write a synoptic sentence or two after the two multiple-choice questions. In general, these sentences mirrored information available in the multiple-choice questions, and hence were not especially enlightening.

REFERENCES

NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

NEISSER, U., & BELLER, H. K. Searching through word lists. *British Journal of Psychology*, 1965, **56**, 349–358.

STERNBERG, S. The discovery of processing stages: Extensions of Donder's method. In W. G. Koster (Ed.), *Attention and performance II*. Amsterdam: North-Holland Publishing, 1969.

(Received July 23, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 506–513

COMPREHENSION TRAINING AND IDENTIFICATION FOR POOR AND GOOD READERS[1]

ROLLIN STEINER, MORTON WIENER[2]
Clark University
AND WARD CROMER
Wellesley College

The relationship of comprehension training (providing contextual information) to identification ("saying" words) was investigated for 24 good and 24 poor fifth-grade readers. Subjects were tested under conditions of paragraph presentation versus single word presentation, and comprehension training (presenting a story summary with synonyms, prior to reading) versus no comprehension training. If comprehension has any effect, it is evident only for good readers. Results indicate that poor readers show no improvement in identification after being given contextual information. This group of poor readers, in contrast to the group of good readers, also fails to utilize syntactic and contextual cues in the material, and seems to treat words as unrelated items in a series.

Reading can be considered as an "identification" activity, that is, "saying" the words aloud, or it can be considered as a comprehension activity in which the primary goal is meaning. Unfortunately, much of the reading literature fails to define adequately which (or both) of these activities is meant by reading. As Wiener and Cromer (1967) have noted, discriminating between identification and comprehension becomes important if we are (*a*) to know which behaviors to assess as critical in reading (however defined); (*b*) to separate reading problems from language problems, intelligence factors, organic defects, etc.; and (*c*) to assess the relationship between identification and comprehension.

Some reading experts (e.g., Durrell, 1958; Gates, 1947) apparently assume that identification of the individual word is the key to good reading. According to this view, once the child learns to identify all the words he is likely to encounter, comprehension will likely follow automatically. Without special training the reader will eventually group words into critical syntactic units. Thus, it is assumed that identification and comprehension are highly correlated.

On the other hand, some investigators (Fries, 1963; Gleason, 1965; Lefevre, 1964; Oakan, Wiener, & Cromer, 1971; Wiener & Cromer, 1967) argue convincingly that the relationship between identification and comprehension is rather tenuous. While some evidence exists that poor identification may be correlated with poor comprehension—good readers seem to comprehend less well with "poor" visual input—research (e.g., Cromer, 1968) points to a rather minimal relationship between good identification and good comprehension.

If comprehension is considered to be somewhat independent of identification, it is necessary to examine the identification process to determine in what ways misidentifying can take place, both in single word and multiple word situations. Kempler and Wiener (1963) view identification as an example of perceptual behavior in which the printed materials are scanned and "part cues" or "partial information" are obtained which become the basis for discrimination between the various graphic shapes (letters) and from which words, phrases, and sentences are elaborated. In this model, all perception is thought of as a "responding to partial information with the particular response being some function of previously learned co-occurrence probabilities [Kempler & Wiener, 1963, p. 355]." This paradigm views perceptual response characteristics as a function of the part cues processed and the response availabilities of the subject. Thus, as the "cue" situations change in the graphic materials, the response patterns evoked also change.

---

[1] This investigation was supported in part by Grant Number OEG-1-9-080043-0005 (010), Project Number 8-A-043, from the United States Office of Education, Department of Health, Education, and Welfare.

[2] Requests for reprints should be sent to Morton Wiener, Department of Psychology, Clark University, Worcester, Massachusetts 01610.

Since all but the most elementary readers appear to use only part of the available information from the printed page, the probability arises of occasional misidentification as a function of (*a*) failure to discriminate the cues; (*b*) failure to utilize "sufficient" cues; (*c*) inappropriate elaboration of cues; (*d*) any combination of the above. For example, a reader might identify the word BAT as if it were RAT, as a function of faulty cue discrimination (mistaking the B for an R). A reader who fails to utilize sufficient cues might identify the word SIX as if it were SOX, the I not being utilized. A reader who elaborates cues inappropriately might well identify the word RECEIPT as if it were the word RECEIVE.

Thus far, cue elaboration for single word situations has been described; however, cue elaboration is also a function of multiple word situations. Cue elaboration should be affected by contextual information. The full range of meanings in a passage feeds back upon cue discrimination and response elaboration. A competent reader is quite unlikely to misidentify BAT as RAT if he has utilized even minimal contextual information which informs him that the passage is about baseball, not about small furry animals. Thus, the discrimination between B and R becomes far less critical. In this way, information derived from context reduces the number of part cues needed to identify correctly a single word or a reading passage.

If the identification process can be conceptualized as some function of both the particular part cues processed and the response availabilities of the reader, a question arises as to the differences between good and poor readers in learned response patterns. Cromer and Wiener (1966) found evidence to support the idea of idiosyncratic response patterns among poor readers, that is, some poor readers appear to elaborate cues differently than do matched good readers. Without inferring any pathology notion, they argued that poor readers either have failed to learn consensual response patterns or learned idiosyncratic response patterns too well. They found that poor readers gave more idiosyncratic responses in several different experimental situations; they gave less consensual responses on word association tasks and on a Cloze procedure in which words are randomly removed from a story and the subject must substitute contextually and syntactically appropriate words.

Apparently, this type of poor reader either has not sufficiently acquired the signaling patterns inherent in the syntax of the language or for some reason fails to make use of these patterns,[3] for these readers typically provide less syntactically appropriate substitutions. Furthermore, such poor readers appear to benefit less from contextual cues. It seems as if the poor reader is forced to make a response in the Cloze procedure from a vastly greater word repertoire than for the good reader. On the other hand, the good reader is able to delimit the response possibilities, utilizing both syntactic cues (e.g., he provides a verb-class form instead of a noun-class form) and contextual cues (e.g., he knows that the story is about hunting, and so seems to limit his verb choice to that context).

In the scanning process which takes place in identification, the question of organization of cues arises. Since the reader cannot take in all cues and process them adequately without slowing his reading speed to a "crawl," he must utilize a system which enables him to correctly identify graphic materials with the fewest possible cues. As mentioned earlier, the identification process is also dependent on the use of meaning cues. It is the meaning or contextual cues which reduce the response possibilities, thus necessitating fewer part cues for accurate identification.

One of the primary techniques for retaining this stream of meaning is "chunking," or grouping multiple cues into single, larger units (phrases, clauses). In terms of information processing, the more "bits" (words) a reader can fit into each chunk, the easier he can hold the reduced number of chunks in storage for processing. In addition, with no grouping of words into larger units, the poor reader is more likely to lose the feedback advantage of context and meaning cues because meaning is carried primarily by phrase units and not by individual lexical items (words).

---

[3] Fries (1963) conceives of reading as responding as adequately to the full range of "language signals" in the graphic mode as in the auditory mode.

The present study, then, is concerned with exploring various possible relationships between comprehension and identification. Wiener and Cromer (1967) point out four different ways in which comprehension (language skills and knowledge of context) potentially aids the identification process: first, knowledge of language structure decreases the possible number of appropriate response possibilities by limiting the response to a given form class or function group; second, the context of a passage delimits the response possibilities (when the reader knows the passage is about "hunting," he can eliminate other response areas); third, knowledge of language structure and context can greatly increase reading speed by decreasing the number of informational units which need to be processed; fourth, knowledge of structure and context provides a possible source of feedback for correcting errors, that is, the reader may experience incongruity between what he hears himself say and what he has already understood to be necessary to fit the relevant context and structure of the given passage being read.

A primary question raised by this study is whether poor readers, who do not appear to use contextual cues, can be aided in the identification process by giving them supplementary contextual information. It has been demonstrated that certain types of poor readers do not organize reading materials into efficient syntactic groupings, and thus lose many essential language cues that aid both the identification and the comprehension process. Furthermore, in failing to utilize these cues, poor readers do not comprehend on a level with good readers. In failing to comprehend as they move through a passage, poor readers must lose much of the context of the reading passage. Losing these contextual cues, they are at a still greater disadvantage in the identification process, for contextual information can greatly reduce the number of cues required for correct identification.

In summary, if the problem for poor readers is primarily a failure to extract cues, it is expected that giving them supplementary contextual information (comprehension training) should lower their identification error rates. Good readers given contextual information would be expected to show only marginal reduction in identification error rates since it is assumed that they already utilize contextual cues as they read.

It is hypothesized that giving contextual information would enable the reader to correct more of his identification errors spontaneously, utilizing context cues as a feedback mechanism. Finally, the study examines the question of the effect of contextual information on a single word mode of presentation and a paragraph mode of presentation.

METHOD

To test these questions, a good reader group and a poor reader group were chosen. Each subject read aloud (identification) under four conditions: a single word presentation with no comprehension training; a single word presentation with comprehension training; a paragraph presentation with no comprehension training; and a paragraph presentation with comprehension training.

Subjects

Subjects for the study were 24 poor readers and 24 good readers in the fifth grade of a New England public school system. The poor reader group consisted of 13 boys and 11 girls reading at least 1½ years below grade level as measured by standardized reading tests; the good reader group consisted of 13 boys and 11 girls reading at or above grade level as measured by the same reading tests. A matched-pairs design was used, with each good–poor reader pair matched for school, class, sex, and approximate age and IQ. Reading test scores for poor readers ranged from 2.5 to 3.5, the mean being 3.28. Reading scores for good readers ranged from 5.0 to 8.1, the mean being 6.0. Ages for both groups ranged from 10 years 4 months to 12 years 3 months, the mean for both groups being 10 years 11 months. IQ scores for both groups ranged from 90 to 110, the mean for poor readers being 97.08 and for good readers 99.20.

Subjects were tested individually in their schools. Good readers averaged 25 minutes for testing, and poor readers 40 minutes. Anxiety was minimized by telling all subjects that this was part of a project to obtain information which could be used to write better textbooks.

Materials and Procedures

Four stories from fifth-grade supplementary textbooks were used. Two forms were employed: a regular paragraph form on 8½ × 11 inch paper; a "word-by-word" form (single word) typed on adding machine paper. This second form of the stories was placed on a machine which the subject "unrolled," with only one word at a time visible.

The four stories were comparable in length (101 to 104 words). For each story a summary was used to provide comprehension training. The summaries were written with minimum repetition of words to avoid prior identification. On comprehension trials, subjects first heard the summary, then read the story. Stories were recorded and analyzed by separate persons for identification errors.

All subjects read the same four stories, each story being read under one of the four experimental conditions.

Each story was paired equally with each condition, and sequential effects were balanced. All poor readers read under exactly the same conditions (sequence and story-condition pairing) as their matched good readers.

After a short orientation by the experimenter, each child was given the following directions:

> I would like you to read four stories out loud for me. Each one is pretty short, and you can take your time. With two of the stories, I'm going to tell you something about the story ahead of time. That way you'll have a good idea of what the story is about before you read it. And two of the stories you're going to read from this machine. It works like this: (Child is shown how machine works). Now, with this first story,... (directions at this point varied, depending on the condition under which the child was reading).

Directions for different conditions:

> *Condition 1:* With this story, I'm just going to have you read it from a regular piece of paper, like you would at school. You're going to find words like...(same words as those repeated from the summary used for comprehension training were given). Just read it as well as you can.
>
> *Condition 2:* With this story, I'm going to have you read it from the machine. I've shown you how the words come up one at a time when you turn the crank as fast or as slow as you like. (You're going to find words like...). Read this story out loud for me, as well as you can.
>
> *Condition 3:* With this story, I'm going to tell you ahead of time what is in the story so you'll know what the story is about (summary of story is read to the child). Now I'm going to read it once more for you, so you have a good idea of what the story is about before you read it (summary is read a second time). Now I want you to read it from this sheet of paper. Just read it as well as you can.
>
> *Condition 4:* (Directions for this condition are a combination of the first part of condition 3, and the first part of condition 2, dealing with the comprehension training and reading by machine.)

All errors not spontaneously corrected by the subject were corrected by the experimenter after the subject had either moved ahead at least two words past the source of the error or hesitated 10 seconds on a word. This was to prevent sequential effects from false output by the subject; it was also an attempt to control for repetition of the error when the same word was encountered later in the story.

RESULTS

The data consisted of the identification errors for each subject under each of the four conditions, analyzed in three categories: initial errors; corrected errors (errors spontaneously corrected by subject); and uncorrected errors (initial minus corrected errors). To control for error rate, the corrected errors were scored as the number of corrected errors over the total (initial) errors. These ratio scores were then transformed to normalize the data. Analyses of

TABLE 1

ANALYSIS OF VARIANCE OF INITIAL IDENTIFICATION ERRORS FOR GOOD AND POOR READERS UNDER CONDITIONS OF COMPREHENSION VERSUS NO COMPREHENSION TRAINING AND PARAGRAPH VERSUS SINGLE WORD MODES OF PRESENTATION

Source of variation	*df*	*MS*	*F*
Between subjects	23	51.99	
Within subjects	168	30.87	
Good vs. poor readers (L)	1	2867.52	69.32*
Error	23	41.37	
Comprehension vs. no comprehension (T)	1	1.02	.09
Error	23	11.67	
Paragraph vs. single word (P)	1	168.75	18.64*
Error	23	9.05	
L × T	1	27.00	4.13**
Error	23	6.54	
L × P	1	11.02	1.16
Error	23	9.43	
T × P	1	46.02	6.80*
Error	23	6.76	
L × T × P	1	6.75	1.45
Error	23	4.94	
Total	191	33.41	

* $p < .05$.
** $p < .06$.

variance for repeated measures were carried out for each of the three error categories.

Table 1 lists results of the analysis of variance for the initial errors. The poor readers made significantly more errors than the good readers ($F = 69.32$, $df = 1/23$). Modes of presentation was significant ($F = 18.64$, $df = 1/23$), more errors being made in the paragraph mode than in the single word mode. All significant means are summarized in the following display. The Level × Training interaction gave significance value at the .06 level ($F = 4.13$, $df = 1/23$). Error rates for good readers increased with comprehension training, with no significant change for poor readers. Also, the Training × Presentation interaction was significant ($F = 6.80$, $df = 1/23$). With the paragraph mode, error rate rose with comprehension training; conversely, with the single word mode, error rate dropped with comprehension training. The effect for comprehension training was not significant ($F = .09$, $df = 1/23$). Neither the Level × Presentation interaction ($F = 1.16$, $df = 1/23$) nor the Level × Training × Presentation interaction ($F = 1.45$, $df = 1/23$) was significant.

List of Mean Number of Identification Errors for the Conditions Found Significant in the Analyses of Variance for Initial, Corrected, and Uncorrected Categories

Initial Errors	
I. Good Readers = 2.59	Poor Readers = 10.32
II. Single Word = 5.52	Paragraph = 7.40
III. Good Readers	Poor Readers
Comprehension = 3.04	Comprehension = 10.02
No Comprehension = 2.15	No Comprehension = 10.63
IV. Single Word	Paragraph
Comprehension = 5.10	Comprehension = 7.96
No Comprehension = 5.94	No Comprehension = 6.83
Corrected Errors	
I. Comprehension = .89	No Comprehension = .59
II. Good Readers	Poor Readers
Comprehension = 1.11	Comprehension = .67
No Comprehension = .57	No Comprehension = .61
Uncorrected Errors	
I. Good Readers = 1.77	Poor Readers = 8.90
II. Single Word = 4.53	Paragraph = 6.14
III. Single Word	Paragraph
Comprehension = 4.02	Comprehension = 6.46
No Comprehension = 5.04	No Comprehension = 5.81

In the category of corrected errors, comprehension training resulted in significantly more spontaneous correction of errors ($F = 14.73$, $df = 1/23$, $p < .01$). The Level × Training interaction was also significant ($F = 11.26$, $df = 1/23$, $p < .01$). In sum, the correction rate for good readers rises significantly with comprehension training; however, for poor readers comprehension training has no significant effect on correction rate.

Neither level (good versus poor readers; $F = 1.85$, $df = 1/23$) nor presentation (paragraph versus single word; $F = .00$, $df = 1/23$) was significant. Finally, the Level × Presentation interaction ($F = .10$, $df = 1/23$), the Training × Presentation interaction ($F = .82$, $df = 1/23$), and the Level × Training × Presentation interaction ($F = .33$, $df = 1/23$) were not significant.

Poor readers made significantly more uncorrected errors than good readers ($F = 63.45$, $df = 1/23$, $p < .001$). Paragraph mode resulted in significantly more errors than single word mode ($F = 18.45$, $df = 1/23$, $p < .01$). Finally, the Training × Presentation interaction was significant ($F = 4.88$, $df = 1/23$, $p < .05$). With the paragraph mode, error rate rises with comprehension training; conversely, with the single word mode, error rate drops with comprehension training.

There was no significant difference attributed to training (comprehension versus no comprehension; $F = .19$, $df = 1/23$). The Level × Training interaction ($F = .76$, $df = 1/23$), the Level × Presentation interaction ($F = 1.70$, $df = 1/23$), and the

Level × Training × Presentation interaction ($F = 1.22$, $df = 1/23$) were not significant.

In summary, good readers perform considerably above poor readers in the identification process, both in initial error rate and in uncorrected (final) error rate. Interestingly, good readers make significantly more identification errors when given comprehension training, with poor readers showing no significant effect. Both good and poor readers make significantly fewer errors with a single word mode than with a paragraph mode of presentation. Finally, comprehension training seems to aid the identification process with the single word mode, but seems to hinder (increases the error rate) the identification process with the paragraph mode.

DISCUSSION

The results of this study do not support the notion that these poor readers fail to use contextual information simply because they cannot pick out contextual cues during the reading process. Rather, the results indicate a more fundamental problem for some poor readers. It appears that they not only fail to extract contextual cues essential for identification, but also fail to utilize such cues in identification even when presented with them. They seem to be identifying words as if the words were unrelated items unaffected by syntactical or contextual relationships.

These results suggest that poor readers, at least in their reading activity, do not make use of various signal systems in their language. Cromer (1968) has demonstrated that poor readers (a "difference" group) who possess adequate intelligence and vocabulary skills do not comprehend on a level with good readers apparently because they fail to organize reading materials into critical syntactic groupings. The present study shows that they fail in utilizing contextual as well as syntactic cues—at least in identifying. Further studies are suggested to analyze the exact types of errors commonly made by good and poor readers to try to specify possible difficulties for the poor readers. It seems safe to conclude that reading difficulties for these kinds of poor readers, at least, encompass the entire range of language signaling devices, and not just problems of identifying individual lexical items.

It was hypothesized that good readers will show some improvement in identification error rates when given supplementary contextual information. Since it is assumed that good readers typically extract contextual cues as they read, supplementary context cues for these readers should be largely redundant. However, the results show an increase in "initial" error rate for good readers when given comprehension training. Thus, good readers respond significantly to comprehension training, but seemingly in the wrong direction.

These good readers appeared to ignore some of the part cues they ordinarily would have needed, perhaps because they had already been given contextual information. However, the errors made by good readers were typically good or anticipation errors, that is, errors that made syntactical and contextual sense. Furthermore, these good readers were able to go back and spontaneously correct many of their initial errors. Apparently, they often noted an incongruity between what they said and what they already understood in the text.

These anticipation errors by good readers were very noticeable in the word-by-word reading. Many of the good readers identified whole phrases before they had cranked the machine far enough to view all the words in the phrase. Then, as the printed words finally appeared, they would spontaneously correct many of their earlier errors. Interestingly, the large majority of good readers grouped their words into meaningful syntactic units, even when reading from the machine which presented one word at a time. They imposed syntactic organization on single word materials either by anticipating the next words or by remaining silent until a phrase grouping had been processed, then identifying it as a single unit.

In contrast, the poor readers failed to utilize even the experimenter's correction of a word when that same word appeared later in the story. Poor readers appeared to be responding to these words as isolated, unrelated items in a series.

In both initial and uncorrected error categories, comprehension training aided identification in the single word mode; however, comprehension hindered identification in the paragraph mode. Differential comprehension training effects for single word versus paragraph modes might be accounted for in terms of the scanning characteristics of these two modes. A single word mode encourages very complete and relatively slow scanning of cues, and discourages chunking into units. A paragraph mode encourages more rapid and less complete scanning, and "invites" chunking. Apparently, comprehension training aids only the slow, complete scanning, where the identification task is made primary by destroying the grouping of words through a single word presentation. Where chunking and comprehending are encouraged, as in paragraph materials, identification suffers with comprehension training. Scanning becomes redundant after comprehension training has provided the meaning of the story.

Throughout all the error categories, significantly fewer errors were made in the single word mode. Both good and poor readers did better when reading word-by-word. In terms of the Kempler-Wiener (1963) model of perception, this finding can be explained on the basis of the degree of completeness of the part cues available. In the paragraph mode, the possibilities for error, given the same response availabilities, are greater than with single word mode, since the reader is scanning and processing a far greater number of units in a given time span. In the single word mode, the part cues are virtually complete, that is, relatively all the graphic representations (letters) are processed, leaving little ambiguity, and thus little choice in terms of response possibilities.

The implications for both the teaching of reading and remedial reading are that identification skills are not sufficient for the reading process. In fact, emphasis on identification skills in the teaching of reading may encourage an "identification set" in which the proper saying of the word takes precedence over critical skills involved in utilizing the full range of contextual and syntactic signaling patterns embedded in graphic language. Previous work by Wiener and Cromer (1967) and Oakan, Wiener, and Cromer (1971) suggests similar conclusions. A large class of poor readers fails to organize reading materials into critical groupings and fails to utilize critical contextual cues needed for adequate identification and comprehension.

A central question concerning reading problems seems to be this: Why should a native speaker have trouble with reading when he has already mastered the full range of complex signaling systems and semantic usages inherent in spoken language? Learning the graphic shapes in reading would appear to be a simple task compared with the formidable job of learning for the first time the complete system of signaling devices in a new language at age 2 or 3. In fact, the assumption by many of direct transfer from auditory to graphic language may lie at the root of the problem. Linguists like Fries (1963) and Gleason (1965) and the Wiener and Cromer (1967) model of reading point out numerous, critical ways in which auditory language differs from graphic language. Speech utilizes patterns of stress, pause, and intonation as essential and primary structure guides. Graphic language, at least as it is widely taught, offers no equivalent devices for signaling syntactic and certain semantic functions—functions critical to comprehension.

Spoken and written language, then, differ in crucial ways, and many issues that come under the rubric of "reading problems" may not be resolved until there is recognition of (and techniques for implementing) the "missing" structure signals in graphic language. Current research is addressing itself to certain of these questions.

REFERENCES

CROMER, W. *The effects of preorganizing reading material on two types of poor readers.* (Doctoral dissertation, Clark University) Ann Arbor Mich.: University Microfilms, 1968, No. 68–17223.

CROMER, W. The difference model: A new explanation for some reading difficulties. *Journal of Educational Psychology,* 1970, **61,** 471–483.

CROMER, W., & WIENER, M. Idiosyncratic response patterns among good and poor readers. *Journal of Consulting Psychology,* 1966, **30,** 1–10.

DURRELL, D. D. First-grade reading success study:

A summary. *Journal of Education,* 1958, **140,** 2–6.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart and Winston, 1950.

FRIES, C. C. *Linguistics and reading.* New York: Holt, Rinehart and Winston, 1963.

GATES, A. I. *The improvement of reading.* New York: The Macmillan Company, 1947.

GLEASON, H. A. *Linguistics and English grammar.* New York: Holt, Rinehart and Winston, 1965.

KEMPLER, B., & WIENER, M. Personality and perception in the recognition threshold paradigm. *Psychological Review,* 1963, **70,** 349–356.

LEFEVRE, C. A. *Linguistics and the teaching of reading.* New York: McGraw-Hill, 1964.

OAKAN, R. The relationship between identification and comprehension skills among good and poor beginning readers. Unpublished master's thesis, Clark University, 1969.

OAKAN, R., WIENER, M., & CROMER, W. Identification, organization, and reading comprehension for good and poor readers. *Journal of Educational Psychology,* 1971, **62,** 71–78.

WIENER, M., & CROMER, W. Reading and reading difficulty: A conceptual analysis. *Harvard Educational Review,* 1967, **37,** 620–643.

(Received July 24, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 514-520

STRUCTURAL COHERENCE IN PICTORIAL AND VERBAL DISPLAYS[1]

RONALD W. SPANGENBERG[2]
Human Resources Research Organization, Fort Knox, Kentucky

The effects of three different levels of structural coherence within verbal and pictorial displays were examined. Six displays were created having identical pictorial or identical verbal items and showing minimal, subgrouped, or overall structural coherence levels. A nonsense syllable was associated with each item as the initial task. The second task was to learn 20 sentences (for which the initial task provides potential mediators) successively presented on a memory drum. Initial learning showed significant superiority of pictorial over verbal display groups. Initial learning of displays showing overall structural coherence provided significant improvement in learning the transfer task as did initial learning of pictorial displays. These results relate both to the design of instructional displays and inferences concerning mental operations.

What effects do different ways of structuring informational displays have on learning? Bruner (1960) says the most basic thing about human memory may be that unless detail is placed in a structured pattern, it is rapidly forgotten. Indeed, as Henle (1966) claims, when the material to be learned does not possess and internal structure, a structure must be imposed on it by the learner. Using paired-associate illustrations, Asch (1969) has shown that when objectively identical contents are differently related, distinctive effects in memory are often produced.

The list structure or organization effects in paired-associate whole list learning is examined for the first time in this study. The question is suggested by combining previous research in paired-associate learning, free recall learning, and word/picture distinctions. The variables used are type of display (two levels: words and pictures) and structural coherence. Structural coherence is the display characteristic describing the degree to which elements of display appear as integrated. Three levels of structural coherence are distinguished:

Minimal structural coherence—No apparent integration of items is cued.

Subgrouped structural coherence—Subgroups of approximately five items each are cued.

Overall structural coherence—The display forms a single unit.

Background material for this research includes material from several fields. Related work in paired-associate learning includes that generated by Epstein, Rock, and Zuckerman (1960), which is now called verbal and pictorial elaboration. It is clearly seen from the results of elaboration studies that the integration of pairs helps learning. Tulving (1962) introduced the concept of subjective organization which has strongly influenced more recent free recall studies which in general hypothesize that increases in the number and integration of items in subjectively organized groups of words accounts for the improvement of performance on repeated free recall trials. These free recall studies show theoretical dependence upon Miller's (1956) unitization theory. Pictures have generally been learned more readily than their labels. Most recently, Wicker (1970) has shown the superiority of pictures to their labels in recognition and recall. Whether organization variables differentially affect words and pictures is a question to be examined in the present study.

Two theoretical assumptions which underlie the present study are:

---

[1] This article is based upon a dissertation submitted to the School of Education, Indiana University, in partial fulfillment of the requirements for the EdD degree. The author gratefully thanks his advisor, Malcolm Fleming, for his help throughout the dissertation project.

[2] Requests for reprints should be sent to Ronald Spangenberg, Human Resources Research Organization, Division No. 2, Fort Knox, Kentucky 40121.

1. The learning of an informational display by an individual is a function of his organizational schema. (An organizational schema refers to the learner's internal representation of an instructional display.)

2. Structural coherence and/or mental elaboration provides relations which may enhance unitization which is a factor in the formulation of an organizational schema. (Unitization refers to a hypothetical mental operation interrelating items into a single psychological unit; mental elaboration refers to a mental operation in which relations between items are generated.)

The procedural modifications of this study are suggested by the work of Reynolds (1966, 1968); that of Bower, Clark, Lesgold, and Winzenz (1969); and Bower and Clark (1969). A comparison is made of display design strategies which provide instances of chunking, overall integration, and minimal integration in both verbal and pictorial displays. The specific question is summarized as: What are the differences in the relative efficiency of initial learning and of learning a transfer task provided by different levels of structural coherence within informational displays of verbal and pictorial materials?

METHOD

Subjects

There were 84 subjects drawn from a group of solicited Indiana University volunteers. All subjects were assigned to one of the seven experimental conditions according to predetermined random order as they appeared individually for the experiment.

Procedures and Materials

One control group and six experimental groups, 12 members per group, were examined. A two-stage transfer learning paradigm with a constant Stage II task was used.

Prefamiliarization. Prefamiliarization with all nonsense syllables was provided by having each subject pronounce each syllable.

Stage I Task. The six experimental treatments will be discussed subsequently.

Stage II Task. The Stage II Task was to learn 20 simple sentences. Each sentence began with the three-letter initials of a hypothetical person and ended with a statement of his occupation. The sentences were:

KOT is a pilot.
NEB is a gas station attendant.
BAF is a shopkeeper.
PUM is a truckdriver.
TUK is a farmer.
HIG is a fruitpicker.
JOW is a teacher.
LEF is an engineer.
QIZ is an announcer.
XES is a minister.
DOS is a policeman.
RAZ is a cook.
FER is a brakeman.
MOV is a bus driver.
YAZ is a stuntman.
SEK is a taxidriver.
CYT is a fireman.
GID is a lifeguard.
WIM is a postman.
VAM is a parking lot attendant.

Initials (consonant-vowel-consonant) of 73.3–93% association value taken from Glaze's (1928) list are used. The sentences constitute factual material about 20 hypothetical persons and may be considered as 20 paired associates. Following each of the Stage I treatments (described below), subjects in all groups were seated before a Stowe memory drum, set in a black screen, and were instructed to learn the 20 sentences. Three learning trials were administered. On each trial, the 20 sentences were presented at a 2-second rate, one at a time, in the drum window. After each trial, the drum was stopped and the subject was given 140 seconds to write on a test sheet as many of the sentences as he could recall. The test sheet was then returned and in 15 seconds a new trial began. The sentences were presented in three successively randomly varied orders to reduce serial order learning effects, and subjects were instructed to write sentences in any desired order on each recall test.

Stage I Treatments. The seven conditions to which the subjects were assigned were:

1. Overall Pictorial Structural Coherence Group. The purpose of this treatment was to impose a pictorial framework for the 20 specified nonsense syllable components related to the Stage II Task. A map providing maximum opportunity for multiple associations was provided. This level of structural coherence does not cue a single grouping strategy, as it produces an overall grouping of the display. The 8½ × 11 inch array appears in Figure 1.

Each subject was instructed to study the initials in the array for 45 seconds in order to reproduce them in their proper positions on a later recall test. After the given 45 seconds, the map was removed and each subject was given 120 seconds to recall and write the missing initials in their correct location on a test map. This procedure was repeated until the criterion of one perfect reproduction was reached.

2. Subgrouped Pictorial Structural Coherence Group. The purpose of this treatment was to impose a categorically structured representation of the stimulus components related to the Stage II learning task. The display consisted of four inte-

grated five-element groups from the map used by the Overall Pictorial Structural Coherence Group. Each element was linked with the specified initials. This level of structural coherence cues a single grouping strategy. The 8½ × 11 inch array appears in Figure 1.

Each subject was instructed to study the display for 45 seconds in order to reproduce the initials in their proper positions on a later recall test. At the end of 45 seconds the display was removed, and the subject was given 120 seconds to recall and write the missing initials in their correct location on a test display.

3. Minimal Pictorial Structural Coherence Group. The purpose of this treatment was to provide no apparent grouping for the pictorial and nonsense syllable elements, identical to those employed with the other two pictorial structural coherence groups. Each combined verbal-pictorial element was bordered. This 8½ × 11 inch array appears in Figure 1.

Each subject was instructed to study the initials for 45 seconds in order to reproduce them next to the appropriate pictorial elements in a 120-second recall test after the study period. This procedure was followed until criterion was reached.

4. Overall Verbal Structural Coherence Group. The purpose of this treatment was to display 20 specified nonsense syllable components related to the Stage II Task in an integrated framework in associated with potential verbal mediators. A story was used to provide overall integration. The story was:

> TUK grew the corn in his field. PUM drove it in his truck past the orchard from which HIG watched. CYT's fire engine, followed by DOS's patrol car passed him. The bus of MOV was hit when FER, from the caboose, was not able to stop the locomotive driven by LEF. "The fire spread to XES's church and JOW's school" announced QIZ from his studio. The truck stopped at NEB's gas station behind WIM's postal van and the taxicab of SEK. As they watched the sky, YAJ parachuted in mid air from the airplane flow by KOT. He missed the parking lot of VAM and was pulled from out of the pool by GID. After the road was cleared the corn was delivered to BAF's shop. RAZ bought it and cooked it in his restaurant for you.

Each subject was instructed to study the story for 45 seconds in order to reproduce the initials in

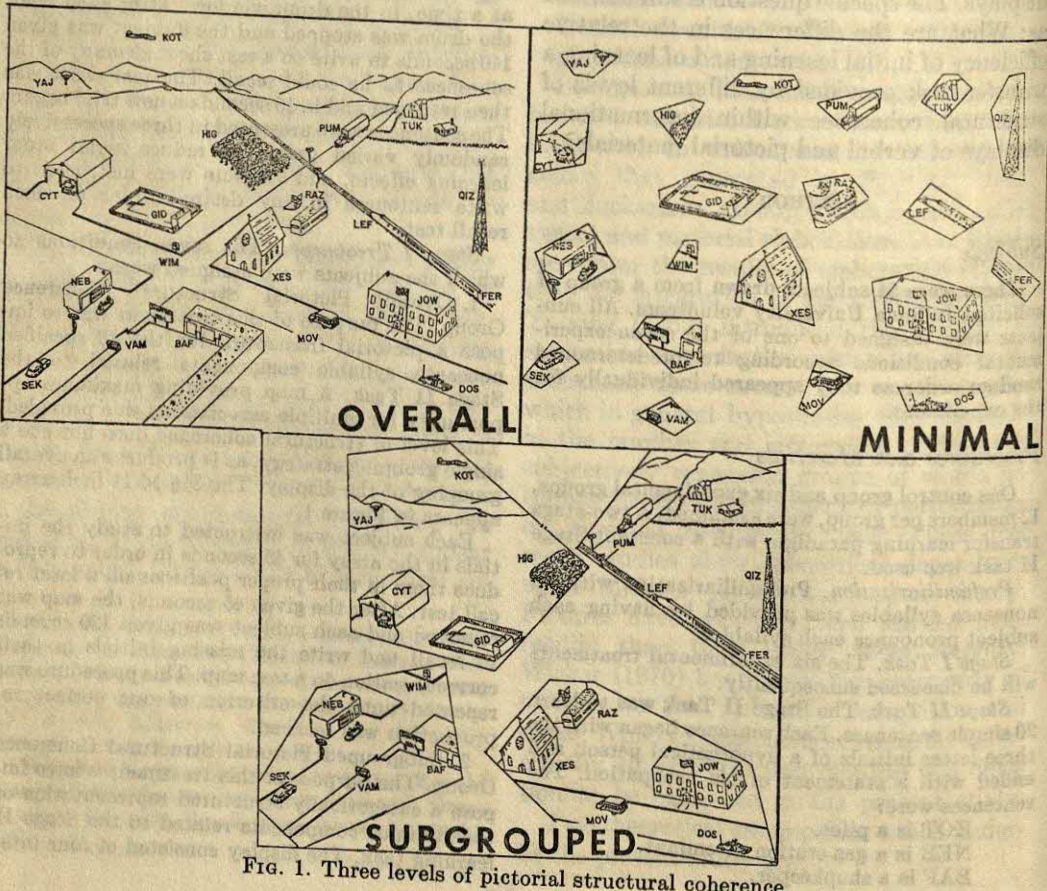


FIG. 1. Three levels of pictorial structural coherence.

their appropriate position. A 120-second test was administered. This procedure was repeated until criterion was reached.

5. Subgrouped Verbal Structural Coherence Group. The purpose of this treatment was to impose a grouping of the 20 specified nonsense syllable components related to the Stage II Task, in a hierarchy in association with the potential verbal mediators. The hierarchy used was a three-level branch hierarchy with Job Locations at the top level. The second level consisted of Local Vehicles, Distance Vehicles, Buildings, and Outdoors. The bottom level consisted of the appropriate initials contiguous with the appropriate verbal items. KOT next to airplane, TUK next to field, and so on.

Each subject was instructed to study the initials on this array for 45 seconds in order to reproduce them in their proper positions on a later recall test. After the given 45 seconds, the hierarchy was removed and the subject was given 120 seconds to recall and write the missing initials in their correct location on a test hierarchy. After the first test, the procedure was repeated until criterion was reached.

6. Minimal Verbal Structural Coherence Group. The purpose of this treatment was to provide an arbitrary list (no inherent organization) of identical nonsense syllables with the same potential verbal mediators as those used in the other verbal structural coherence groups. The minimal structural coherence group was provided a list of pairs. These pairs consisted of the verbal items, after which appeared the appropriate nonsense syllables, such as KOT next to airplane, and so on.

Each subject was instructed to study the initials for 45 seconds in order to pair them with the appropriate word in a 120-second test which followed. This procedure was followed until criterion was reached.

7. Rest Group. A control group learned the list of sentences in the Stage II Task, but received no Stage I treatment. This control provides a baseline learning curve for the Stage II Task.

Dependent Variables and Design

The dependent variables were: (*a*) sum of correct responses, first three trials initial learning, (*b*) total number of trials to criterion in initial learning, and (*c*) sum of correct responses, first three trials of Stage II Task.

The design was Type of Display Level by Structural Coherence Level by Subject (nested in Type of Display and Structural Coherence Levels).

RESULTS

The data were analyzed by employing the computer program BMD 08V for analysis of variance (Dixon, 1968), and where appropriate, the Tukey test (Winer, 1962).

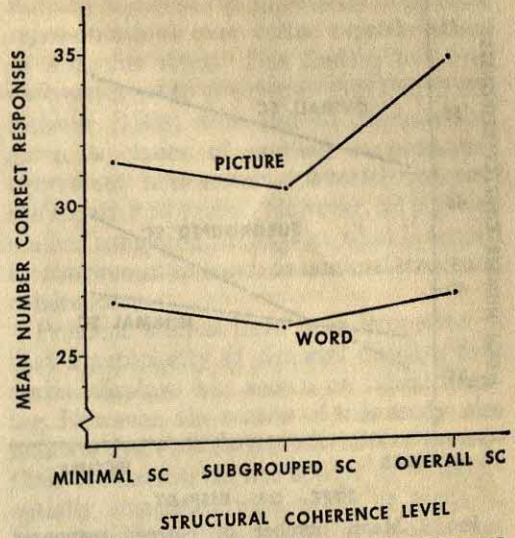


FIG. 2. Mean number of correct responses summed over the first three trials of Stage I learning for two types of displays across three levels of structural coherence.

For initial (Stage I) learning, the analysis of variance showed that the type of display effect over three trials was significant. The groups who were shown pictorial materials scored higher than the groups who were shown verbal materials (Figure 2, Table 1). This type of display effect accounted for 18.6% of the variance (following Hays, 1966). There was no significant interaction. The total number of trials to criterion in initial learning, when analyzed, indicated no significant differences.

An analysis of the Stage II (transfer) Task showed two significant main effects and no significant interaction (Figure 3, Table 1). There were significant differences between groups from different levels of structural coherence, accounting for 18.7% of the variance. A Tukey test indicated

TABLE 1

ANALYSIS OF VARIANCE SUMMARY

Source	*df*	First three trials Stage I		First three trials Stage II	
		MS	*F*	*MS*	*F*
Type of display (D)	1	583.68	5.66*	618.35	7.09**
Structural coherence (SC)	2	42.35	.41	870.50	9.98***
D × SC	2	20.59	.20	57.39	.66
S (D × SC)	66	103.09			
Total	71				

* $p < .025$.
** $p < .01$.
*** $p < .005$.

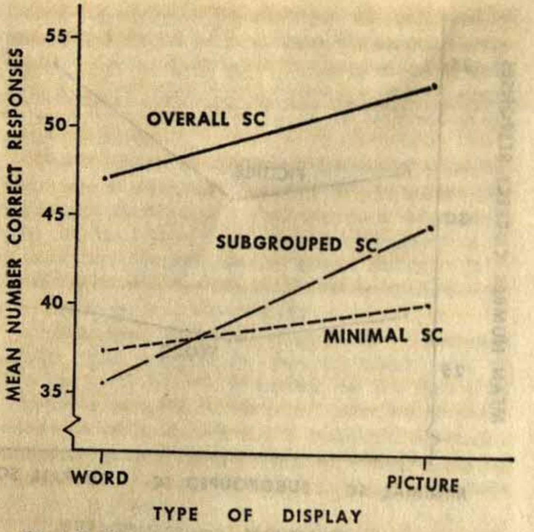


FIG. 3. Mean number of correct responses summed over the first three trials of Stage II learning for three levels of structural coherence across two types of display.

that the overall structural coherence groups were significantly superior to both other groups (which were not significantly different). Pictorial display groups were reliably superior to verbal display groups (accounting for 6.3% of the variance). All initial learning tasks resulted in significant difference ($p < .01$) from the control (rest) group on the transfer task using the Tukey test.

DISCUSSION

This research demonstrates a learning consequence attributable to the apparent degree of integration of an initial learning display. Earlier studies showing that pictorial materials are easier than others to learn are verified, and transfer of learning attributable to pictorial materials is shown.

This study defines structural coherence as a display characteristic which describes the degree to which elements of a display appear to be integrated. An overall level of structural coherence was shown to provide a significant positive increment of learning in a verbal transfer task after the initial learning of displays that exhibit different levels of pictorial or verbal structural coherence. However, the efficiency of the initial learning was not shown to be affected by the structural coherence level of the Stage I display.

There are two characteristics in the design of this experiment which are used together for the first time. The first is that the Stage I learning of all displays was to the same criterion of one perfect trial. The second is that all stimulus items on all displays appeared simultaneously. All Stage I learning conditions, therefore presented a whole, within which three conditions suggesting three degrees of structural coherence were presented.

The results suggest that overall unitization by the learner provides him a more stable organizational schema. Reynolds (1968), using only eight-item pictorial displays, suggests that it is the cuing for integration of elements, not the elements themselves, which facilitates transfer. The present study verifies and extends his conclusion. Reynolds (1968) found transfer only for his integrated pictorial (overall structural coherence) display group. The small numerical size of Reynolds' displays probably contributed to this difference in results. An overall unitization is provided by the popular mnemonic devices which use mental elaboration to generate relations.

The results of this study show that subgrouped structural coherence level displays did not demonstrate as much transfer as overall structural coherence level displays. Displays with a subgrouped structural coherence (five items each) did not enhance transfer when compared to displays with minimal structural coherence levels. Mandler (1968) created a speculative hierarchical model of the mental organization of verbal memory. This model suggests the efficiency of a subgrouping strategy. It was inferred from extending the implications of Mandler's model that, when using integrated subgroups of approximately five items each, superior transfer might result. The present results do not make it possible to extend Mandler's model to displays of the size and nature used in this study. The distinction must be made that Mandler used displays of from 50–100 familiar words, and was theorizing about the structure according to which familiar words would be stored in memory.

The results of this study may be related to a dual-component memory hypothesis formulated by Slamecka (1968) for free recall learning. Slamecka suggested that during the presentation of a free recall list, the learner encodes and stores a general representation of the list structure, the specific item traces being stored independently at the same time. Later, the learner uses the general representation of the list to form a retrieval plan which guides his search for the item traces. The present study assured the storage of all item traces by requiring the learning of the Stage I displays to the criterion of one perfect reproduction. The Stage I displays, however, may have cued differing retrieval strategies which could also be used in the transfer task. This can be interpreted as a transfer of structure. What is learned in the Stage I Task may be not only the initials, but also the structure which may be used to retrieve them.

The results of this study support Asch's (1969) conclusion that objectively identical contents, when differently related, produce distinctive effects in memory. His conclusion was based primarily on the relating of paired versus unpaired items. The present study extends the findings of distinctive effects of organization to displays of up to 20 items (or pairs of items). It also shows a particular family of relations, those called structural coherence, as a significant source of variance in transfer of learning (and by implication, the original quality of learning).

An obvious implication of this research is that the learning of displays of up to 20 items should be designed to emphasize the overall integration of the items to be learned. The perceptual feature of showing the interrelatedness of each item to the whole seems to enhance the quality of learning. Artificial breakdowns of subgroups of items (not having a self-integrity explicitly directed toward the overall integration of the display) appears to be a poor design strategy. Also, certain instructor-created groupings and sequences may be detrimental to the highest quality of learning.

Wimer and Lambert (1959) reported that nonsense syllables were more readily learned when objects, rather than printed names of the objects, were the stimulus items. When comparing pictures and words, nonsense syllables were attached to pictures more readily than to words. Lumsdaine (1949) initially reported the superiority of pictorial representations over verbal representations as stimulus items. This finding has been authenticated to the degree that Gagné and Rohwer (1969) were able to conclude that given a choice of method in presenting equivalent information, pictorial materials are superior to verbal. However, all previous studies compared displays at what is termed by the present research as minimal structural coherence.

Previous studies have been supported in that a superiority of pictorial displays over verbal displays was shown on initial learning. However, the results of this study also indicate that elaborative activities or perceptions of relations do not appear to be differentially associated with words or pictures. No significant interaction or trend was observed between Type of Display Level and Structural Coherence Level.

The transfer task was a verbal reproduction of statements for which the Stage I learning stimulus items functioned as potential mediators. A significant positive increment in learning the transfer task was attributable to the pictorial potential mediators. It appears possible to assume that pictorial materials are not only more efficiently learned, but also provide a more stable organizational schema.

REFERENCES

ASCH, S. A reformulation of the problem of association. *American Psychologist*, 1969, **24**, 92–102.

BOWER, G. H., & CLARK, M. C. Narrative stories as mediators for serial learning. *Psychonomic Science*, 1969, **14**, 181–182.

BOWER, G. H., CLARK, M. C., LESGOLD, A. M., & WINZENZ, D. Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 323–343.

BRUNER, J. S. *Process of education*. New York: Vintage Books, 1960.

DIXON, W. J. (Ed.) *BMD biomedical computer programs*. Berkeley & Los Angeles: University of California Press, 1968.

EPSTEIN, W., ROCK, I., & ZUCKERMAN, C. B. Meaning and familiarity in associative learning. *Psychological Monographs*, 1960, **74**(4, Whole No. 491).

GAGNÉ, R. M. & ROHWER, W. D., JR. Instructional psychology. *1969 Annual Review of Psychology*, 1969, **20**, 381–418.

GLAZE, J. A. The association value of nonsense syllables. *Journal of Genetic Psychology*, 1928, 35, 255–269.

HAYS, W. L. *Statistics for psychologists*. Chicago: Holt, Rinehart & Winston, 1966.

HENLE, M. Cognitive skills. In J. Bruner (Ed.), *Learning about learning*. Washington, D.C.: United State Government Printing Office, 1966.

LUMSDAINE, A. A. Ease of learning with pictorial and verbal symbols. Unpublished doctoral dissertation, Stanford University, 1949.

MANDLER, G. Association and organization: Facts, fancies, and theories. In T. Dixon & D. Horton (Eds.), *Verbal behavior and general behavior theories*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

MILLER, G. A. Human memory and the storage of information. *IRE Transactions on Information Theory*, 1956, 2, 129–137.

REYNOLDS, J. H. Cognitive transfer in verbal learning. *Journal of Educational Psychology*, 1966, 57, 382–388.

REYNOLDS, J. H. Cognitive transfer in verbal learning. II: Transfer effects after prefamiliarization with integrated versus partially integrated verbal-perceptual structures. *Journal of Educational Psychology*, 1968, 59, 133–138.

SLAMECKA, N. J. Examination of trace storage in free recall. *Journal of Experimental Psychology*, 1968, 76, 504–513,

TULVING, E. Subjective organization in free recall of "unrelated words." *Psychological Review*, 1962, 69, 344–354.

WICKER, F. W. On the locus of picture-word differences in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 52–57.

WIMER, C. C., & LAMBERT, W. E. The differential effects of word and object stimuli on the learning of paired-associates. *Journal of Experimental Psychology*, 1959, 57, 31–36.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received August 10, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 521–525

EFFECT OF SENTENCE-COMBINING EXERCISES UPON SEVERAL RESTRICTED WRITTEN COMPOSITION TASKS[1]

M. R. VITALE, F. J. KING,[2] D. W. SHONTZ, AND G. M. HUNTLEY
Florida State University

The performance of Negro fifth-graders who engaged in a drill on a series of simple and complex sentence combining exercises either as observers, as active learners with observers present, or as individual learners, was compared with that of controls who were drilled individually on complex exercises only, or on no exercises. The results showed that the experimentals obtained significantly more correct responses than the controls on the complex exercises; while the observers and individual learners used a greater number of words/T-unit than the active learners or the two controls on composition tasks requiring the rewriting of a short segment of prose. In addition, a significant correlation was found between words/T-unit and the frequency with which the embeddings practiced in the exercises were used on the composition tasks.

The importance of language skills as prerequisites to many kinds of classroom learning can hardly be overemphasized. Yet, as Bandura and Harris (1966) have argued, research in psycholinguistics has been undertaken primarily to categorize the usage and development of language behavior under naturalistic conditions rather than to identify conditions governing language acquisition and modification. The purpose of the present investigation was to demonstrate that language behaviors dependent upon knowledge of grammatical sentence structure were amenable to experimental manipulation. More specifically, the investigation was conducted to determine the effect of several series of sentence-combining exercises similar to those used by Mellon (1967) and Ney (1968) upon grammatically related written composition tasks under conditions in which the sentence combining drill was conducted either in an individual or in a modified observational learning paradigm (Bandura, 1965).

METHOD

Subjects

Ten children were selected randomly from a public elementary school population of Negro fifth-graders whose California Test of Mental Maturity scores were within the average range (90–110). The test-score restriction was made in order to help eliminate children with reading difficulties.

Materials

Three kinds of experimental materials were constructed to correspond to three stages of the investigation. The first stage was a learning task in which the subjects drilled upon sentence combining exercises; the second, a test task in which subjects attempted to combine a series of more complex sentence combining exercises; and the third, a rewriting (composition) task in which subjects rewrote short segments of prose.

The learning materials consisted of 120 exercises. Each learning exercise required a different pair of simple declarative sentences to be combined using a single grammatical embedding. Forty of the exercises required a factive embedding, 40 required a relative embedding, and 40 required a reduced-relative embedding. Within each category of 40 exercises, 20 required a verb-phrase embedding and 20 required a noun-phrase embedding. Examples of a verb-phrase embedding and of a noun-phrase embedding for each category appear in Table 1. For each exercise, the pair of sentences to be combined was typed on one side of a 4 × 6 inch unlined card and the same two sentences, along with the appropriate combined sentence, on the reverse side of the card.

---

[1] The authors gratefully acknowledge the assistance and cooperation of the principal, Herman Landers, of Lincoln Elementary School in Tallahassee, Florida, and of teachers, Helen Jones and Aldorse Harris, during the time this study was conducted. The statistical analyses were made possible in part by National Science Foundation Grant GP-5144 to the Florida State University Computation Center.

[2] Requests for reprints should be sent to F. J. King, Division of Instructional Research and Service, Florida State University, Tallahassee, Florida 32306.

TABLE 1

EXAMPLES OF A NOUN-PHRASE AND VERB-PHRASE EMBEDDING OF EACH CATEGORY OF LEARNING EXERCISE

Category of embedding	Noun-Phrase	Verb-phrase
Reduced-relative	The bee dropped the honey. The bee was small.	The rain wet the men. The men were tired.
	The small bee dropped the honey.	The rain wet the tired men.
Relative	The teacher opened the window. The teacher was tired.	The boy had a ticket. The ticket was for the show.
	The teacher that was tired opened the window.	The boy had a ticket which was for the show.
Factive	Something pleased me. The sun was shining.	We knew something. The puppy broke the lamp.
	That the sun was shining pleased me.	We knew that the puppy broke the lamp.

The test materials consisted of eight exercises. Each test exercise required a set of four sentences to be combined into a single sentence using the three embeddings practiced separately in the learning task. One of the test exercises appears in the following display. Each test exercise was typed on a 4 × 6 inch unlined card using a format similar to that of the learning exercises. The order in which the four sentences to be combined appeared on the card indicated the proper sequence of the three embeddings. The following is one of the test exercises used in the study:

1. The woman called for help.
2. The woman believed something.
3. The rabbit ate the plant.
4. The rabbit was white.

Answer: The woman who believed that the white rabbit ate the plant called for help.

The writing materials consisted of two 60-word passages. Each passage contained 12 simple declarative sentences similar in form to those used in the learning and test exercises. Thus, the sentences within each passage could be combined using factive, relative, and reduced-relative embeddings. One of these passages appears below.

> The boy is ten years old. The boy has a sister. The sister is young. The sister knows something. The boy likes something. The sister plays baseball in the field. The field is dirty. His mother helps the boy with his homework. The homework is hard. The boy is active. His mother hopes something. The boy will study in school.

In addition to these, two shorter passages, each containing four sentences and each identical in grammatical form to the test exercises, were included as separate writing tasks. For use in the experiment, each passage was mimeographed on a separate page.

Design

Two subjects were randomly assigned to each of the treatment conditions indicated in Table 2. The four subjects assigned the paired condition were divided into two groups each containing a learner and an observer. The observer in each pair watched the learner work through the learning material after which both the learner and observer individually attempted the test material. The two subjects assigned the individual condition separately worked through first the learning material and then the test material. The remaining subjects served as controls; those in the first control group attempted individually only the test material, and those in the second control group attempted neither the learning nor the test material. Writing tasks were administered to all 10 subjects.

Procedure

The study was conducted over 12 consecutive school days. The subjects' progress through the learning and test material was partially equated by

TABLE 2

DESIGN FOR THE PRESENT STUDY

Treatment group	Learning task	Test task	Writing task
Paired learner	Yes	Yes	Yes
Paired observer	Yes (observe only)	Yes	Yes
Individual	Yes	Yes	Yes
Control 1	No	Yes	Yes
Control 2	No	No	Yes

varying their daily work time. The learning materials were administered on Days 1 through 8; the test material on Days 6 through 10; and the writing materials on Day 12.

The subjects assigned the learning material always attempted first the 40 reduced-relative exercises, then the 40 factive exercises, and finally the 40 relative exercises. Within each category the 40 exercises were presented in a random order with the constraint that Exercises 1–10 required a verb-phrase embedding and Exercises 11–20 required a noun-phrase embedding. In the administration of the learning exercises, subjects were required to respond correctly to a series of five exercises before new exercises were presented.

The subjects assigned the test material attempted four of the eight exercises one day and the remaining four the following day. The exercises assigned on a particular day were attempted four times in a constant order.

In the writing task, each subject first rewrote each of the two shorter passages and then one of the two longer passages. The order in which the two shorter passages were attempted was counterbalanced over the treatment conditions as was the assignment of the longer passage attempted by each subject.

A small quiet room was used for the administration of the learning and test material. The experimenter and the subject attempting the exercises sat at right angles to each other. For each individually presented learning and test exercise, the experimenter placed a card on the table in front of the subject showing the sentences to be combined. After reading the sentences silently, the subject guessed aloud how the sentences were combined on the reverse side of the card. The experimenter then turned the card over and the subject read aloud the combined sentences. For the paired subjects, the procedure was identical to that for individual administration except that the observer was seated next to the learner and that the observer rather than the learner read aloud the correct combined sentence after the learner's guess. Throughout the administration of the material, the experimenter stressed that subjects were to "guess the sentence on the reverse side of the card." For the two subjects in the first control group who attempted the test material but not the learning material, special care was taken to be certain that the task was understood.

The writing task was administered to the group of 10 subjects in the school auditorium. The instructions presented orally by the experimenter stressed that the subjects were to "rewrite each story telling it in a better way." At no time was it suggested that subjects should combine the sentences to improve the story.

RESULTS

Although the subjects' performance on the writing task served as the primary data in the study, a separate analysis of performance upon the test task was also conducted.

TABLE 3

NUMBER OF CORRECT TEST EXERCISES FOR THE GROUPS ASSIGNED TEST MATERIAL

Treatment group	Frequency of correct responses[a]
Paired learner	10
Paired observer	17
Individual	11
Control	0

[a] $\chi^2 = 15.89$, $df = 3$, $p < .01$.

A one-sample chi-square test was used to determine whether the frequency of correct test responses differed among the four treatment groups. The obtained $\chi^2 = 15.89$, $df = 3$, was significant at the .01 level. In inspecting the number of correct responses for each of the treatment groups shown in Table 3, it should be noted that none of the control subjects obtained a correct response.

Two separate analyses of subjects' performance on the writing task were undertaken. A one-sample chi-square test was employed to determine whether the subjects under the five treatment conditions differed in the total number of reduced-relative, relative, and factive embeddings used in rewriting the three passages. The resulting $\chi^2 = 34.68$, $df = 4$, was significant at the .01 level. The number of embeddings used by each treatment group appears in Table 4. It should be noted that none of the subjects in

TABLE 4

MEAN W/T AND FREQUENCY OF REDUCED-RELATIVE, RELATIVE, AND FACTIVE EMBEDDINGS ON THE THREE WRITING TASKS FOR EACH OF THE TREATMENT GROUPS

Treatment group	Frequency of reduced-relative, relative, and factive embeddings[a]	Mean W/T unit[b]
Paired learner	2	4.883
Paired observer	12	10.193
Individual	13	9.433
Control 1	0	4.780
Control 2	0	4.620

[a] $\chi^2 = 34.68$, $df = 4$, $p < .01$.
[b] $F = 42.00$. $df = 4/5$, $p < .001$.

either control group used an experimental embedding in rewriting the three passages.

The dependent variable used in the second analysis was the mean words/T-unit (W/T) for each of the three passages rewritten by subjects. Hunt (1965) has defined a T-unit as one main clause plus whatever subordinate clauses are attached to or embedded within that main clause. The W/T statistic has been shown (Hunt, 1965) to be an objective and valid definition of syntactic maturity in writing. Analysis of variance for the 5 × 3 factorial design in which repeated observations were taken across the three writing passages for each of the five treatment groups showed a significant effect of groups ($F = 42.00$, $df = 4/5$, $p = .001$). The effects of passages ($F < 1$) and of the Passage × Group interaction ($F < 1$) were not significant. The mean W/T for each treatment group appears in Table 4. Duncan's multiple-range test found the mean W/T for the individual group and the mean W/T for the paired observer group to be significantly greater ($p < .001$) than the W/T means for the paired learner group and for the two control groups. No other differences between means approached significance.

To estimate the extent to which T-unit length was dependent upon use of the three kinds of embeddings practiced in the learning material, the Pearson r between W/T and (number of) embeddings/T-unit (E/T) was computed for the subjects' overall performance on the writing task. The obtained r was significant ($r = .82$, $df = 8$, $p < .05$).

DISCUSSION

The differences in performance on the writing task among the five treatment groups as indexed by the W/T statistic illustrated the importance of exposure to the sentence combining exercises which composed the learning material. However, due to differences in performance between the paired observer and the individual learner groups compared to the paired learner group (shown in Table 4), no general statement can be made about the reliability of the effect of exposure to the learning material independently of the conditions under which it occurred. Unlike the paired learner group, which on the writing task performed at the same level as the two control groups not exposed to the learning material, the paired observer and individual learner groups did use the kinds of embeddings upon which they were drilled in their writing task.

In considering possible interpretations of these results, it must be stressed that the observational learning paradigm used in the present study was a modification of that defined by Bandura (1965). In the Bandura paradigm the observer is given the opportunity to perform the modeled behaviors only after the observational portion of the experiment is completed. In the present study, the observer performed part of the learning task during the acquisition period. In fact, the paired learner and observer were each assigned a part of the learning task which the individual learners performed entirely. The paired learner only guessed the correct response, whereas the paired observer read the correctly combined exercises aloud. Thus, it could be argued that if the paired observers first covertly attempted to combine each pair of sentences and then read aloud the correct response, their task was identical to that of the individual learners. Further it could be argued that the task of the paired learners who were only required to guess the correct response was identical both to that of the paired observers and the individual learners, with the important exception that the manner of providing informational feedback was a grossly inefficient one.

In demonstrating the validity of the W/T statistic as a definition of syntactic maturity in writing, Hunt (1965, p. 157) called attention to the potential usefulness of a sentence building program in accelerating the writing skills of elementary and secondary school students. Recently, working with seventh-grade subjects, Mellon (1967) compared the effect of a sentence-combining program (in conjunction with classroom instruction in a minimal transformational grammar) with a conventional parsing treatment and with a control conditon. Mellon found that compared to the two other groups, the experimental subjects experienced more than one year of additional growth in writing maturity as defined by Hunt's T-unit statistics. It is important, however, that the present study along with the work reported by Ney (1967) implies that a program of the type suggested by Hunt may be applied fruitfully to children in primary grades without concurrent classroom instruction in grammar. Consistent with this notion are arguments offered by O'Donnell, Griffin, and Norris (1967) and Hunt (1968) arising from transformational grammatical theory stating that T-unit growth as a valid index of syntactic maturity directly reflects the number of E/T. The effect of sentence combining drill upon T-unit length and the correlation of .82 between W/T and E/T found in the present study lend empirical support to the above views.

REFERENCES

BANDURA, A. Vicarious processes: A case of no-trial learning. In L. Berkowitz (Ed.), *Advances in exprimental social psychology*, Vol. II. New York: Academic Press, 1965.

BANDURA, A., & HARRIS, M. B. Modification of syntactic style. *Journal of Experimental Child Psychology*, 1966, **4**, 341–352.

HUNT, K. W. *Grammatical structures written at three grade levels*. Champaign, Ill.: National Council of Teachers of English, 1965.

HUNT, K. W. An instrument to measure syntactic maturity. Florida State University, 1968. (Mimeo)

MELLON, J. C. Transformational sentence-combining: A method for enhancing the development of syntactic fluency in english composition. (Project 5-8418, Cooperative Research Program, Office of Education, United States Department of Health, and Welfare) Cambridge: Harvard University Press, 1967.

NEY, J. W. Conditioning syntactic performance of children at varying grade levels by audio-lingual drills on transformations. Paper presented at the annual meeting of the American Educational Research Association, February, Chicago, 1968.

O'DONNELL, R. C., GRIFFIN, W. J., & NORRIS, R. C. *Syntax of kindergarten and elmentary school children: A transformational analysis*. Champaign, Ill.: National Council of Teachers of English, 1967.

(Received August 12, 1970)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 526–530

IMAGERY AND SENTENCE LEARNING[1]

RICHARD C. ANDERSON[2] AND JANET L. HIDDE
University of Illinois

Twenty-four subjects, who were told they were helping norm materials for a future experiment, either rated the pronunciability or the image-evoking value of 30 sentences. After one presentation of the sentences, a surprise test was administered in which the subject attempted to recall the verb and object of each sentence given the subject noun as a retrieval cue. Subjects who rated imagery recalled over three times as many words as those who rated pronunciability. It is argued that imagery instructions facilitated learning by causing subjects to process the sentences in a meaningful fashion.

The role of imagery in paired-associate learning has been under intensive investigation recently. It has, for instance, been demonstrated that the "image-evoking value" of words correlates highly with the learnability of these words, more highly than any other characteristic that has been investigated including meaningfulness, Thorndike-Lorge frequency, and semantic differential ratings (Paivio, 1969). Other research indicates that instructions to create mental images from words strongly facilitates paired-associate learning (Bower, in press; Paivio & Yuille, 1967).

The purpose of the experiment described here was to see whether imagery instructions facilitate sentence learning. This is not a foregone conclusion. According to one view, which can be labeled the *product* hypothesis, associative learning is accomplished when the learner possesses a mediator (the "product") to link a word pair. It can be argued that imagery instructions (and verbal mediation instructions) cause subjects to search for mediators, which they otherwise would not consistently do. When a mediator is provided by placing a noun pair in a sentence (cf. Rohwer, 1967), no further facilitation from imagery instructions would be expected, if the product hypothesis is correct. An alternative notion can be called the *process* hypothesis. According to this view what is important is the process of conceiving a link between the things and events named by the words, rather than the link itself. To put this another way, it is the act of bringing to mind meaningful representations that is the precursor to learning. On the basis of the process hypothesis, imagery instructions would be expected to facilitate sentence learning because such instructions induce the subject to meaningfully process the words whereas he may otherwise do no more than translate them into speech.

METHOD

Subjects

The subjects were 24 student volunteers, paid $1.00 for their participation, and randomly assigned to groups.

Sentences

Thirty sentences were constructed according to the following pattern:

The NOUN$_1$ VERB the NOUN$_2$.

The verbs were transitive, and all were in the past tense. The nouns were names of animals or kinds of people, selected at random from the animate nouns with a frequency of from 200 to 700 occurrences per 4.5 million in the Thorndike-Lorge (1944) List T. The verbs were chosen from the same frequency range. The mean frequencies of subjects, verbs and objects were 366, 370, and 378, respectively.

The sentences were constructed by colligating two nouns and a verb at random. The resulting sentence was included in the list if it made sense, provided there was not an obvious association between the nouns or between the verb and either noun. No two words within a sentence were allowed to start with the same letter. Some examples of the sentences are: The lady examined the robin; The butcher kicked the stranger; The queen greeted the professor.

---

[1] The research reported in this paper was supported in part by the Advanced Research Projects Agency through the Office of Naval Research under Contract ONR Nonr 3985(08).

[2] Requests for reprints should be sent to Richard C. Anderson, who is now at the Department of Education, University of Leeds, Leeds L52 9JT, England.

Design and Procedure

Subjects were told that they were helping norm sentences for use in a future experiment. Half the subjects received imagery instructions and half received pronunciability instructions. Subjects in the imagery group were asked to form an image of the event described in each sentence and then rate the vividness of that image. Subjects in the pronunciability group were instructed to repeat the sentence aloud three times and then rate its pronunciability.

Each sentence was presented on a 3 × 5 inch file card for 7 seconds. This interval was signaled by the words "Begin" and "Stop" heard from a tape recorder. At the signal "Begin," the experimenter held up a card displaying a sentence. When "Stop" was heard, the experimenter removed the card and the subject rated imagery or pronunciability during the 3-second interitem interval on a sheet provided for this purpose. The order of sentences was randomized for each subject by shuffling the set of cards upon which the sentences were typed.

A surprise recall test was given after one presentation of the sentences. The test consisted of a mimeographed list of the subject nouns. Following each noun was a blank into which the subject was asked to write the rest of the sentence that contained that noun. The instructions encouraged recall of sentences the subject was not sure of and indicated that the subject should report a part of a sentence even when the remainder could not be recalled. Two different test orders were prepared and randomly assigned to the subjects within groups. The test was subject paced.

RESULTS

Verbs and objects were scored separately. The Imagery group recalled a mean of 10.6 verbs while the Pronunciability group averaged only 3.3 ($t = 5.16$, $df = 22$, $p < .01$). For objects, the recall means were, respectively, 10.4 and 3.0 ($t = 5.20$, $df = 22$, $p < .01$).

Table 1 shows the mean frequency of types of recall errors. Most errors, especially in the Pronunciability group, were errors of omission. Subjects in the Imagery group sometimes substituted a synonym for the correct word (e.g., *baby* was reported instead of *child*). These errors were labeled "synonym intrusions." Most "other intrusions" were intralist intrusions.

TABLE 1

MEAN FREQUENCY OF TYPES OF RECALL ERRORS

Treatment group	Error		
	Omission	Synonym intrusion	Other intrusion
Imagery			
Verb	15.1	1.2	3.2
Object	15.4	1.5	2.7
Pronunciability			
Verb	24.8	0.0	1.9
Object	25.7	.1	1.3

Despite the fact that the instructions emphasized that the subject should report sentences and parts of sentences even when he was not sure of accuracy, sentences tended to be recalled correctly as a whole or not at all. For the Imagery group, the probability of recalling an object, $P(O)$, was .34, whereas the conditional probability of recalling an object given that the verb in the same sentence had been recalled, $P(O \mid V)$, was .78. For the Pronunciability group, $P(O)$ was .10 while $P(O \mid V)$ was .69. The same relationship was observed for the recall of verbs. For the Imagery group, $P(V)$ was .36 and $P(V \mid O)$ was .81. The figures were .11 and .73 for the Pronunciability group.

Figure 1 indicates that there was a strong recency effect. The later a sentence appeared the more likely it was to be recalled. Figure 2 plots the relationship of Imagery ratings and sentence recall for the Imagery group. Sentences which evoked vivid images were better recalled than sentences which evoked vague images, but the effect was not strong. This is puzzling in the light of the robust relationship between rated imagery and pair learning (Paivio, 1969). Pronunciability ratings were unrelated to recall.

DISCUSSION

The chief finding of the present study was that people instructed to form images of the events described in sentences recalled more than three times as many words on a surprise test as people who merely pronounced the sentences. In a paired-associate study which also involved incidental learning, Bower (in press) found that a group instructed to form images recalled twice as many words as a group that pronounced each pair. Imagery instructions might facilitate paired-associate learning because the instructions induce the subject to search for

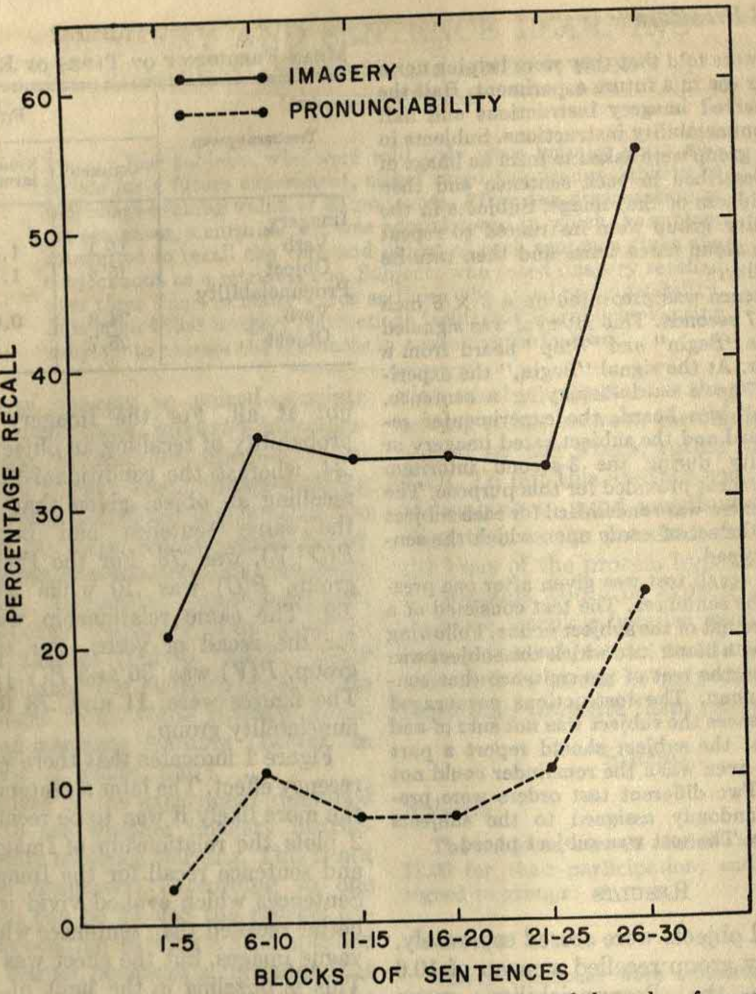


Fig. 1. Percentage of verbs and objects recalled as a function of the order of presentation of sentences.

mediators. However, such an explanation cannot account for the improvement that imagery instructions cause in sentence learning since presumably a sentence provides a mediator.

The fact that imagery instructions facilitate sentence learning is consistent with a process hypothesis (cf. Anderson, 1970). According to this view, there are several stages or levels of processing required for verbal learning. Two of these stages are auditory encoding and semantic encoding. Auditory encoding is translating words into speech while semantic encoding is representing words in some meaningful form. It is argued that people tend to do no more processing than the task requires. Specifically, the contention is that subjects in the Pronunciability group seldom brought to mind the meaning of the words they were articulating. In fact, the instructions to read the sentences aloud three times probably suppressed spontaneous semantic encoding and occupied the time which otherwise might have been devoted to meaningful processing.

The presumption is that mental imagery is a form of meaningful representation, perhaps a peculiarly efficacious form. If this is the case, it is not surprising that subjects instructed to form mental images learned readily even without an intention to learn. The mature learner who intends to learn

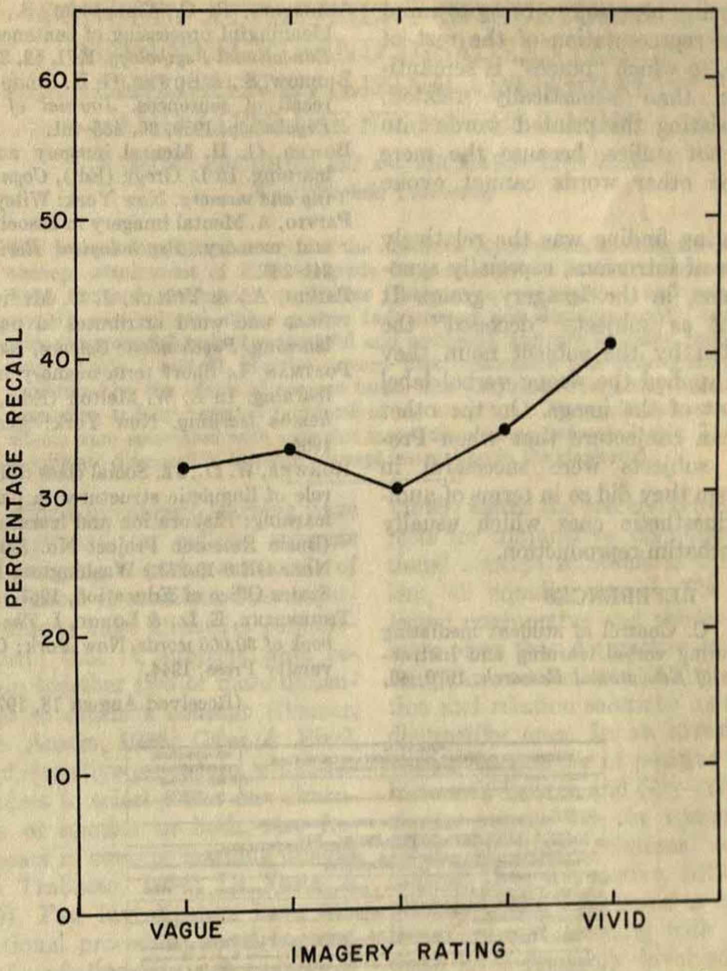


FIG. 2. Percentage recall of verbs and objects in the Imagery group as a function of imagery ratings.

will usually complete the operations required to learn, but an intention to learn is unnecessary if the task itself requires full processing (cf. Postman, 1964).

There are now several lines of evidence to support the utility of a distinction between auditory and semantic encoding, and to show that procedures that force semantic encoding promote learning. Bobrow and Bower (1969) gave instructions to compose a sentence which was a sensible continuation of each sentence that was presented. For instance, if they saw "The farmer discovered a diamond," a subject might respond, "He sold it to a jeweler and used the money to buy a tractor." The subjects in the continuation condition recalled twice as many object nouns as control subjects required to read each sentence aloud three times. To compose a sensible continuing sentence requires comprehension of the presented sentence whereas pronouncing the presented sentence does not. Anderson, Goldberg, and Hidde (1971) constructed sentences in which the last word was determined by the rest of the sentence. Subjects who supplied the missing last word of each sentence as they read learned more rapidly than subjects who read whole sentences. One of the incomplete statements was, "Girls carry lots of junk in their ______." The argument is that to complete the sentence with the word "purses" requires a person to bring to mind a meaningful representation of the rest of the sentence, to which "purses" is semantically rather than acoustically related. Simply translating the printed words into speech will not suffice, because the mere sound of the other words cannot evoke "purses."

An interesting finding was the relatively large number of intrusions, especially synonym intrusions, in the Imagery group. It may be that as subjects "decoded" the images elicited by the subject noun they occasionally applied the wrong verbal label to an element of the image. On the other hand, one can conjecture that when Pronunciability subjects were successful in storing an item they did so in terms of auditory and kinesthetic cues which usually permitted verbatim reproduction.

REFERENCES

ANDERSON, R. C. Control of student mediating processes during verbal learning and instruction. *Review of Educational Research*, 1970, 40, 349-369.

ANDERSON, R. C., GOLDBERG, S., & HIDDE, J. Meaningful processing of sentences. *Journal of Educational Psychology*, 1971, 62, 395-399.

BOBROW, S., & BOWER, G. H. Comprehension and recall of sentences. *Journal of Experimental Psychology*, 1969, 80, 455-461.

BOWER, G. H. Mental imagery and associative learning. In L. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, in press.

PAIVIO, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, 76, 241-263.

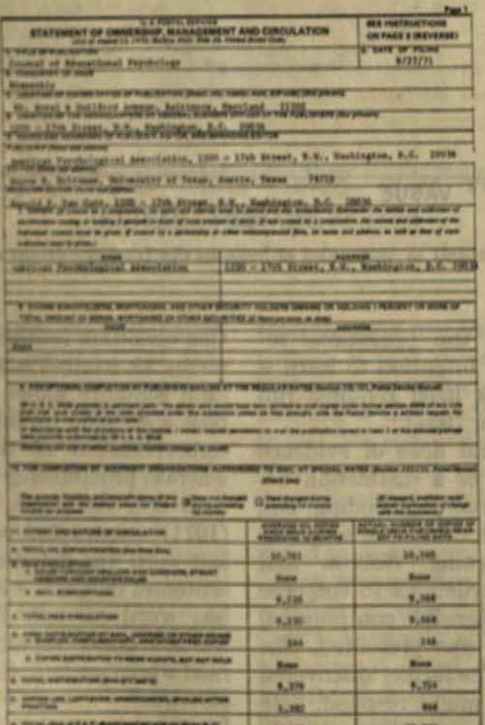
PAIVIO, A., & YUILLE, J. C. Mediation instructions and word attributes in paired-associate learning. *Psychonomic Science*, 1967, 8, 65-66.

POSTMAN, L. Short term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964.

ROHWER, W. D., JR. Social class differences in the role of linguistic structures in paired-associate learning: Elaboration and learning proficiency. (Basic Research Project No. 5-0605, Contract Nonr OE 6-10-273.) Washington, D. C.: United States Office of Education, 1967.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Columbia University Press, 1944.

(Received August 18, 1970)



Page 1

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

SEE INSTRUCTIONS ON PAGE 2 (REVERSE)

Journal of Educational Psychology
1971, Vol. 62, No. 6, 531-538

CONCEPT ATTAINMENT OF CULTURALLY ADVANTAGED AND DISADVANTAGED CHILDREN UTILIZING ARTIFICIAL AND LIFELIKE STIMULUS TASKS

SAMUEL SECURRO, JR.[1] AND RICHARD T. WALLS[2]
West Virginia University

The present experiment compared the effects of experimental treatments on concept attainment of 120 fifth-grade boys and girls. The design crossed three concept attainment conditions (disjunctive, conjunctive, and relational), two cultural status groups (advantaged and disadvantaged), two stimulus material tasks (meaningful and artificial), and two repeated problems (Trials 1 and 2). Results showed that attainment was differentially facilitated by the effects of concept tasks, with conjunctive requiring fewest responses, and by repeated problems of the same type. No significant main effects were associated with stimulus materials or socioeconomic levels. The results are discussed in terms of concept instruction in the classroom.

Research efforts in concept learning have largely concentrated on college subjects sorting and grouping stimulus materials of various geometrical dimensions. The majority of experiments have utilized a conjunctive procedure; that is, subjects were required to join together two or more dimensional values to attain a concept (Bruner, Goodnow, & Austin, 1956; Osler & Fivel, 1961). The disjunctive procedure, which requires a subject to select either one dimensional value or another or both, also frequently appears in concept learning designs (Conant & Trabasso, 1964; Di Vesta & Walls, 1969). Few investigators have utilized a relational procedure, requiring manipulation of a relationship, such as more of one dimensional value than another. Carroll (1964) concludes that many of the concepts learned by children are of a relational character; they deal with relationships among dimensions rather than dealing with combined absence or presence of dimensions.

While no studies with children have reported a comparison of the three concept solution strategies, Hunt and Hovland (1960) tested the preferences of college subjects for conjunctive, disjunctive, and relational concept solutions to the same problem, all equally correct. The subjects selected conjunctive and relational solutions more often than disjunctive ones. Similarly, Laughlin and Jordan (1967) found conjunction and relation solutions more often than disjunctive ones. In an investigation that varied the number of positive and negative instances, Bourne and Guy (1968) obtained similar results for the conjunctive treatment, but found relational solution more difficult than disjunctive. Although several attempts have been made to evaluate relational concept learning with children, the tasks have primarily involved "matching" procedures that required the subjects only to identify one relevant dimension, and subjects did not relate dimensional values as outlined in Bruner et al. (1956).

Heidbreder (1946) obtained more rapid concept learning with more lifelike materials (tree, duck) than with materials featuring artificial (geometric) dimensions. Carroll (1964) states that no absolutely established positive relationship exists between learning a "nonsense concept", such as "BLAB," as the label of a prescribed geometrical figure, and the learning of the word "longitude" in a school setting. Concepts learned in the school depend upon dimensions which in themselves represent difficult concepts.

Some authors have suggested that the disadvantaged child may have greater difficulty dealing with tasks requiring representational or mediational activity (Havighurst, 1966; Jensen, 1966). He has been compared to his middle-class counterpart in intelligence, reading level, aptitude, and, to a lesser extent, concept learning ability. Assessment for the most part has been accomplished through pencil and paper measures, and results on the surface favor the middle-class child. Biber (1970), however, comments that many deficiencies in conceptual functioning have been decided upon seemingly through the use of subjective experiences rather than empirical means.

---

[1] Now at West Virginia State College.

This article is based on a dissertation submitted to the graduate school of West Virginia University in partial fulfillment for the degree of Doctor of Education. The authors express appreciation to Serrafine Tallerico (principal) and to the teachers of Hayes Avenue, Wiley, and Wolfdale Schools.

[2] Requests for reprints should be sent to Richard T. Walls, Department of Educational Psychology, 806 FTA, West Virginia University, Morgantown, West Virginia 26506.

It should thus be noted that, first, the majority of experiments have employed artificial or geometric stimulus materials, and these materials have typically been artificially combined into dimensions to represent a concept such as "two, blue, squares." Few studies have dealt with school related concepts or employed meaningful stimulus materials (Carroll, 1964; Glaser, 1970). Second, there is a disproportionate amount of evidence on conjunctive and disjunctive attainment in comparison to the paucity of evidence on relational concept attainment, and few studies have specifically dealt with relational concepts. Third, research on children's concept attainment has been directed largely toward the middle-class child. Although some concept research has been accomplished with disadvantaged preschool children (Kinz, 1969), the topic has not been systematically explored either with the disadvantaged child in a school setting or across concept solution strategies. If disadvantaged children do indeed have greater difficulty with tasks requiring more representational activity, they may perform more poorly with the geometric stimuli than with the lifelike problems of the present design.

The investigation compares traditional geometric stimulus tasks and lifelike stimuli on conjunctive, disjunctive, and relational concept types solved by culturally disadvantaged and advantaged elementary school children. It is assumed that concept learning should improve across repeated problems of the same concept type. Further, conjunctive attainment should be more efficient than either disjunctive or relational. As noted, conflicting evidence exists with regard to the latter two solution types. The present experiment should also contribute to better understanding of the processes operating to facilitate or hinder attainment in lifelike concept learning tasks (Carroll, 1964; Heidbreder, 1946). Overall, nondisadvantaged subjects should attain the concepts more efficiently than the disadvantaged children.

METHOD

Subjects

The subjects were 120 fifth-grade children (57 girls and 63 boys) from five classes in three Washington County, Pennsylvania, public schools. Although the classes did not contain exactly half disadvantaged and half advantaged children, 60 of each classification could be identified from the five classrooms. Subjects ranged in chronological age from 10 to 11 years.

Students were identified as culturally disadvantaged ($N = 60$) by various combinations of the following: (*a*) parent(s) receiving welfare payment, (*b*) parent(s) engaged in unskilled occupation, (*c*) student receiving free milk or lunch, (*d*) student residing in low income housing section, (*e*) student participated in Head Start or other compensatory education program, (*f*) student did not attend kindergarten classes, and (*g*) teacher, counselor, and administrator knowledge of home conditions. Analogous criteria were the basis for selection of the advantaged subjects.

Design

Half of the subjects in each socioeconomic group were assigned by reference to a table of random numbers to a geometric task condition ($n = 30$) or to a lifelike task condition ($n = 30$). These were further assigned randomly to three concept treatment groups—conjunctive, disjunctive, or relational ($n = 10$). Each subject was exposed to two attainment tasks in his particular concept condition. It was thus a mixed design with a $2 \times 2 \times 3$ factorial format and two repeated measures. To the extent possible, sex assignments were balanced across conditions. Intelligence was not controlled for across the socioeconomic dimension. However, considerable evidence is accruing which questions the proposed relationship between intelligence and concept attainment. Several findings support the assumption that in concept attainment experiments in which the task is explicitly defined, the effects of intelligence disappear (Guthrie, 1970; Osler & Fivel, 1961). Bloom's (1970) eight measures of concept skill yielded no significant correlations with intelligence. In general, these findings indicate that the effects of intelligence are minimized by explicit instruction to subjects.

Stimulus Materials

The stimulus materials were dichotomized into lifelike and geometric stimuli. For the lifelike stimuli, dimensions were contrived to be representative of meaningful elements recognizable from the subjects' sensory environment. Artificial tasks featured a geometric form (circle) in variations of size, color, and number.

There were 18 individual card instances duplicated to form a deck of 36, 5 × 8 white flexible cards for each of the two stimulus materials treatments. The geometric task cards contained circles on all card instances, making three relevant dimensions: color (red and blue), size (large and small), and number (1, 2, and 3). The 18 instances were so constructed that one, two, or three large circles appeared at the top of the card, and on the bottom of each card were one, two, or three small circles. For example, a card instance in the geometric deck had "three, large, red" circles at the top and at the bottom of the card, "two, small, blue" circles spaced equally in a horizontal pattern. The large and small circles were 1⅜ and 11/32 inches in diameter, respectively. It should be noted that the value large was always at the top of the card; all circles at the top of a given card were the same color while all circles at the bottom of the card were the other color. This construction was necessary to adequately assess the three concept types according to the specifications of Bruner et al. (1956).

The lifelike stimuli were presented on cards similar to the laboratory deck, but employed lifelike stimulus dimensions in a more meaningful array, adapted from Bruner et al. (1956). The three dimensions were size (adult and child), affect (happy and sad), and number (1, 2, and 3). As in the geometric deck, individual card instances featured various combinations of these dimensional values, constructed in the same way. For example, a card instance from the school related deck featured at the top, "three, adults, smiling," and at the bottom, "three, children, frowning," (see Figure 1).

It should be noted that in experiments making such comparisons, there is a necessity to control the amount of information available to subjects exposed to different experimental conditions (Sechrest & Wallace, 1962). The design employed here attempted to parallel the structure of the lifelike and the artificial deck cards by the use of both large and small figures, by varied numerical combinations from one through three, and by featuring in both decks a third dimension with two values, such as two values for affect (happy and sad). It should also be noted that the dimensions values of the naturalistic stimuli might tend to create a small amount of information that could be perceived by subjects as irrelevant; for example, in the adult value stimuli, elderly adults were featured as well as younger adults. However, Osler and Fivel (1961) reported that the use of such materials did not increase the complexity of the task significantly for children in high and average ability groups.

In selecting concept tasks, numbers of positive and negative instances were controlled, to the extent possible. In the deck of 36 cards, positive and negative instances for conjunctive tasks numbered 12 and 24, respectively, for relational, 12 and 24 also, and for disjunctive, 20 and 16.

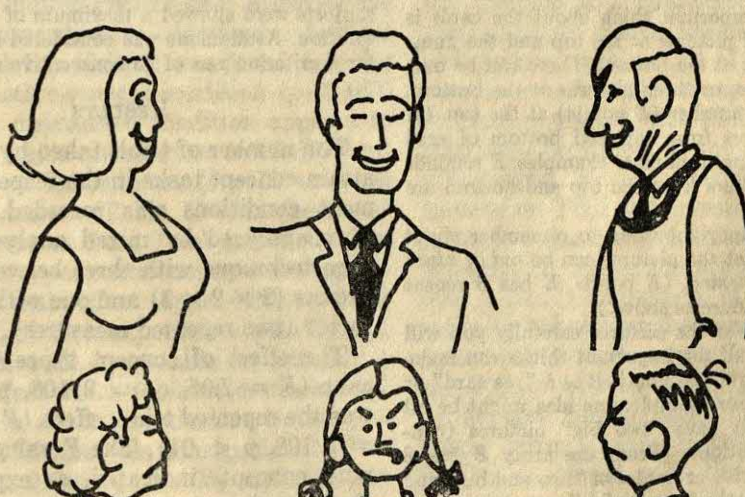


FIG. 1. Sample card instance from the lifelike deck.

Procedure

Each subject was tested individually by an experimenter in a room equipped with chairs and a table. The subject sat opposite the experimenter, and the appropriate stimulus cards were placed on top of the desk. Also, on the desk top was a 5 × 8 inch chalk-drawn rectangular "box" labeled "no-box." Both the "no" and the "yes" boxes were within easy reach of, and equidistant from the subject. The subject was informed that he would be playing a game in which he was to determine which cards go in the yes-box (positive instances) and which in the no-box (negative instance of the same concept).

Each subject was required to attain two problems appropriate to the assigned treatment condition. For example, in the lifelike conditions, one conjunctive problem was "three *and* smiling"; one disjunctive problem was "three smiling *or* frowning"; and for relational, "more smiling than frowning." Subjects in the appropriate geometric stimulus conditions were required to attain parallel problems.

Before proceeding further, the experimenter arrayed nine representative card instances from the deck before a subject and instructed:

> First of all the cards have pictures at the top and pictures at the bottom of the cards. All the pictures on the cards are *circles, people*. (Italics indicate appropriate word(s) choice.) The first important thing about the cards is that the size of the pictures can be big or small, (*E* points to several examples). Now you point to all the big pictures. Good! All the little pictures. Good! (If *S* omits any, *E* prompts *S* by saying, "do you see any others?").
>
> The next important thing about the cards is the number of pictures at the top and the number of pictures at the bottom. There can be one, (*E* points) two, or three pictures at the bottom, and the same number (*E* points) at the top. (If *S* adds numbers from top and bottom of card when responding to number examples, *E* reminds him that numbers from the top and bottom are separate.)
>
> The last important thing to remember about the cards is that the pictures can be *red* or *blue*; *smiling* or *frowning*, (*E* points. *E* has *S* repeat pointing procedure as above.)
>
> If you look at the pictures carefully you will see that some of the important things can make an idea that will tell you if it is a "yes card" or a "no card". For example, one idea might be all the cards that have "two big" pictures (conjunctive condition). (From the array *E* has *S* select all positive examples of "two and big" and place them in the "yes-box"; the remaining are placed in the "no-box," and *E* explains which important things were used. Also *E* explains that this is not the idea that the *S* has to figure out, but one similar to it. Examples are based and constructed appropriately to the concept condition of the *S*.) Another idea might be all the cards that have "two blue" pictures. (*E* repeats above procedure).
>
> Now, I have another idea about the cards, but this time I am not going to tell you the idea. I am going to give you a "yes" and a "no" card to help start you. (*E* places cards appropriately).
>
> Your job is to figure out the idea by looking at each card in the deck and deciding if the important things on each card make it a "yes" or a "no" card. If you think the card is a "yes" place it in the "yes-box" (*E* points), or if you think it is a "no" card, place it in the "no-box." After you put the card in one of the boxes, I will tell you if it was a "yes" or a "no" card. Any questions?

Following this orientation, the subject was required to attain two tasks appropriate to the assigned treatment condition. Attainment problems in all conditions were presented in counterbalanced order to adjust for any inherent differences in difficulty. Basically, the reception paradigm was utilized with a correction procedure; that is, single card instances were presented by the experimenter to a subject to which the subject responded yes or no, was corrected if necessary, and placed the card in the correct pile. Both decks of stimulus cards were initially arranged to allow for yes card instances to be distributed equally throughout the decks. Moreover, to facilitate memory, the last yes and no card after each subject had completed the entire deck remained in view while the experimenter shuffled the deck. The stimulus cards in each deck were duplicated in order to control for the possibility of subjects memorizing the card instances. Identical instructions and procedures were utilized with subjects in both task conditions. The subjects were allowed a maximum of 150 trials for solution. Attainment was considered successful after a criterion run of 15 consecutive correct trials.

RESULTS

The number of trials taken by subjects to attain concept tasks in the respective treatment conditions was recorded. The data were analyzed by mixed analysis of variance technique with three between-subjects factors ($2 \times 2 \times 3$) and one within-subjects factor (two repeated measures).

The effect of concept types was significant ($F = 7.06$, $df = 2/108$, $p < .01$) as was the repeated trials effect ($F = 25.18$, $df = 1/108$, $p < .01$). The F value associated with concepts indicated, as expected, that the mean number of responses for attainment varied significantly across the concept types. Application of Duncan's multiple range test to determine differences among the three concept types, yielded differences

TABLE 1

MEANS AND STANDARD DEVIATIONS OF TRIALS TO CRITERION FOR ALL TREATMENT CONDITIONS

Concept type	Disadvantaged subjects				Advantaged subjects			
	Artificial		Lifelike		Artificial		Lifelike	
	Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2
Relational								
$\bar{X}$	78.00	74.10	68.30	72.10	82.80	52.20	94.50	87.10
SD	45.77	37.80	40.58	40.19	47.00	29.66	48.79	48.04
Conjunctive								
$\bar{X}$	56.00	37.50	62.90	57.10	48.10	41.70	56.20	38.40
SD	31.20	17.32	47.37	38.62	26.70	17.66	44.43	15.86
Disjunctive								
$\bar{X}$	76.70	51.50	101.00	61.10	82.00	56.70	85.70	59.40
SD	38.25	28.75	45.34	38.08	53.00	28.13	34.64	29.75

between conjunctive and disjunctive ($p < .05$), conjunctive and relational ($p < .05$), but no difference between disjunctive and relational. In the conjunctive conditions, subjects required fewer attempts to attain their tasks than the subjects in either the disjunctive or relational conditions.

A significant interaction effect of concept levels with trials was found ($F = 3.33$, $df = 2/108$, $p < .05$). As may be noted in Table 1, while the absolute gains from Trial 1 to Trial 2 were approximately equal for conjunctive and relational, the disjunctive condition showed greater gain. Applying the Duncan procedure, a significant decrease from Trial 1 to Trial 2 was found only for the disjunctive concept condition ($p < .05$). Thus, the disjunctive condition appears to make the major contribution to the Concept Condition × Trials interaction. However, it should be noted (Table 1) that there was an improvement from Trial 1 to Trial 2 in all concept types contributing to the significant trials main effect. Apparently, subjects as a result of their experiences with Trial 1 of a given task, were able to transfer a learning effect which facilitated attainment.

The main effect for stimulus tasks, geometric and lifelike, was nonsignificant ($F = 2.08$, $df = 2/108$, $p > .05$) and did not significantly interact with the effects of status, concept types, or trials. Neither task in the present investigation effected an advantage in attainment behavior. Concept attainment was not differentially affected by the use of the two simulus tasks directly or in combination with any of the aforementioned effects.

No major differences in performance were associated with the two socioeconomic status groups—disadvantaged and advantaged ($F = .02$, $df = 1/108$, $p > .05$). The similarity of these groups is evident in Table 1. Further, status of subjects did not significantly interact with any other factors. Finding no support for the hypothesized differences between the two apparently indicated that performance in categorizing of the relevant and irrelevant stimuli of the present study can be accomplished equally well among such groups of children. However, in examining the mean performance of subjects presented with lifelike tasks in the relational condition, it may be noted that advantaged subjects required more responses on Trial 1 than disadvantaged subjects. In order to check this observation statistically, a Fisher's exact probability test was made. Comparison of the cells was made by tabulating the number of most extreme attainment scores. In this instance, the highest extreme scores occurred in the advantaged lifelike relational cell, and the lowest extreme scores in the disadvantaged lifelike relational cell. The difference between the two frequencies using the Fisher's test was significant at $p < .05$ for a two-tailed test.

Although the effects of concept types and trials were statistically significant, the analysis of variance procedure does not yield estimations of the amount of variance attributable to each factor proportional to the total variance. Accordingly, the Omega square procedure was conducted for the significant main effects and interactions. Having such estimations available is considered essential for indicating the magnitude, as distinct from statistical significance, of variation due to individual effects or interactions (Vaughn & Corballis, 1969). Results showed that the effect of concept types accounted for .30 and trials effect .15 of the total variance. Less than 1% (.044) was attributable to the interaction of concept type with trials. Thus, a relatively stronger effect was associated with the concept types and trials as main effects.

It was noted by the experimenter throughout the experiment, that female subjects, both advantaged and disadvantaged, generally appeared to be attaining their respective tasks somewhat more quickly than the males. Since little has been reported on sex differences in the concept literature, a post hoc treatment of the data was deemed justified. The analysis of this hypothesis was conducted utilizing the original design format, with the exception that all the previous independent measures used were further subdivided by sex.

Although the main effect for sex differences did not reach significance, a trend effect in the expected direction was found ($F = 3.37$, $df = 1/96$, $p < .10$). No significant interactions with sex were obtained.

DISCUSSION

In general, the findings of the investigation indicated that the efficiency and correctness with which subjects attained their respective concepts depended to a great extent on the type of concept and the effect of repeated trials. The findings support the hypothesis that subjects in conjunctive treatment conditions require fewer trials than subjects in either disjunctive or relational conditions.

The significant effect of concept types indicates that children can effectively manipulate bits of information in a conjunctive manner. Although the subjects performed better with conjunctive compared to disjunctive and relational tasks, it should be emphasized that, for the most part, subjects were successful in attainment of the latter two. The fact that individuals do not utilize disjunctive thinking may be an artifact of the traditional concept learning experiment and not that the human brain is not built to think disjunctively. Even though the disjunctive concept has been shown to be more difficult than the conjunctive concept for many age groups, it is reasonable to expect that with training and feedback, individuals can effectively learn strategies for disjunctive solution (Snow & Rabinovitch, 1969).

Generally, it would appear that teachers and textbook writers have adopted a conjunction based mode for explaining and teaching concepts. Such a mode has taken several forms—classifications (categorizing information by some species characteristics), definitions (simply defining the concept verbally or in writing), synonyms (identifying concept words with the same or nearly the same meaning), sentences (using the concept word in a sentence), and a mixed program (Johnson & Stratton, 1966). The use of such methods has been criticized since they expose children only to minimal learning and do not facilitate expanding of initially learned concepts to a more generalized understanding.

Pella (1966) indicates that the teaching of concepts in science primarily by conjunction procedures is unrealistic because many such concepts do not have standard parts or attributes which can be "added up" to equal a constant meaning. Perhaps for classroom instruction too much emphasis has been placed on conjunctive concept types; it may be that concepts from science or other subject-matter areas may better be learned using disjunction or relation based strategies primarily. The problem remains to be systematically explored in future research.

Finding a significant learning effect for subjects in disjunctive conditions indicates that disjunctive concept types can be effectively learned when appropriate training is provided for the learner. Although no learning effects were found for conjunctive and relational tasks, many subjects in conjunctive conditions attained such tasks extremely fast (15–25 trials). Thus a significant improvement in performance on Trial 2 was improbable. For the relational tasks, the small change in performance from Trial 1 to Trial 2 may reflect the greater difficulty of the tasks; the difficult empirical relationships represented in relating activity may be less easily mediated (Griffin, 1967).

The significant main effect of improvement in attainment from Trial 1 to Trial 2 indicates that similarities from previous instruction (Trial 1) arise in current instruction or learning (Trial 2). Such findings have been documented by previous investigators (Keppel & Postman, 1966). Of course, the prior learning effect can be a negative as well as a positive contributor to attainment. In the classroom, once a conceptual position is formed and it is deemed necessary to dissolve such a position, the learner may experience difficulty in overcoming that formation. However, properly sequenced series of exposure to many problems of the same general type have a positive transfer or facilitative effect on human as well as infrahuman learning (Di Vesta & Walls, 1967; Harlow, 1949), and incorporation of different concept types into such series could facilitate broadly based and broadly applicable concept learning.

Although no overall significant performance effects were associated with the stimulus tasks, there was a differential noted in the number of trials to criterion required on Trial 1 for advantaged subjects versus disadvantaged subjects presented with lifelike stimuli in the relational task only. In the complex relational task this finding may indicate that the realism of the lifelike tasks created irrelevant information, and although an orientation to relevant dimensions was provided for subjects initially, such information could have interfered with attainment or increased the difficulty of the lifelike attainment task. Several authors have suggested that less emphasis on incorporation of detail and realism in the lessons may facilitate learning (Bruner, et. al., 1956; Dwyer, 1967).

The status of the subject was not differentially related to the difficulty of the concept tasks, with the exception of the tentative finding on the lifelike relational task discussed previously. Although evidence exists which indicates that such differences are not readily found, a plausible explanation is that the attainment tasks utilized may not have reflected a general response advantage for either group. That is, with explicit instructions such tasks apparently do not provide an advantage because of differences in previous learning experiences at home or in school. In the classroom, disadvantaged youngsters, as a group, may attain concept tasks as efficiently as advantaged children under specific instructions and provision for a learning effect by the teacher. Under these conditions such boys and girls apparently can effectively manipulate bits of information implicitly and demonstrate such a facility behaviorally. Based on the reported sex finding tendency and within the limitations of the subjects selected for the investigation, one would speculate that female fifth-grade subjects may be better able to verbally mediate and categorize information to facilitate attainment. Developmental differences exisiting between female and male subjects may have contributed to the trend found (McCandless, 1961).

The present findings appear to recommend that teachers begin to use disjunctive and relational concept approaches in addition to the commonly employed conjunction presentation strategies for promoting concept learning. Similarly, the prerequisite training necessary for efficient sequential concept learning should be provided. With detailed and specific preliminary instructions regarding the nature of the task teachers should assume that, as a group, disadvantaged children can learn conceptual skills as efficiently as their middle-class counterparts and expect them to perform accordingly.

REFERENCES

BIBER, B. Goals and methods in preschool programs for disadvantaged children. *Children*, 1970, **17**, 15–20.

BLOOM, R. On the cognitive functioning of young children: A theory, instrument, and developmental norms. Paper presented at the American Educational Research Association, Minneapolis, March 1970.

Bourne, L. E., & Guy, D. E. Learning conceptual rules: The role of positive and negative instances. *Journal of Experimental Psychology*, 1968, **77**, 488-494.

Bruner, J., Goodnow, J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.

Carroll, J. Words, concepts, and meanings. *Harvard Educational Review*, 1964, **34**, 178-202.

Conant, M. B., & Trabasso, T. Conjunctive and disjunctive concept formation under equal-information conditions. *Journal of Experimental Psychology*, 1964, **67**, 250-255.

Di Vesta, F. J., & Walls, R. T. Transfer of object-function in problem solving. *American Educational Research Journal*, 1967, **4**, 207-215.

Di Vesta, F. J., & Walls, R. T. Rule and attribute identification in children's attainment of disjunctive and conjunctive concepts. *Journal of Experimental Psychology*, 1969, **80**, 498-504.

Dwyer, F. M., Jr. Adapting visual illustrations for effective learning. *Harvard Educational Review*, 1967, **37**, 250-263.

Glaser, R. *Concept learning and concept teaching.* Learning Research and School Subjects. Eighth Annual Phi Delta Kappa Symposium on Educational Research, 1968.

[illegible], A. F. A philosophical approach to the subject-matter preparation of teachers of history. Unpublished master's thesis, University of Kansas, 1967.

[illegible], J. V. Effects of instruction and socioeconomic status on concept learning in children. Paper presented at the American Educational Research Association, Minneapolis, March 1970.

Harlow, H. The formation of learning sets. *The Psychological Review*, 1949, **56**, 51-65.

Havighurst, R. J. Who are the socially disadvantaged? *Journal of Negro Education*, 1964, **33**, 210-217.

Heidbreder, E. The attainment of concepts: The psychological interpretation problem. *Journal of Genetic Psychology*, 1946, **35**, 191-223.

Hunt, E. B., & Hovland, C. I. Order of consideration of different types of concepts. *Journal of Experimental Psychology*, 1960, **59**, 220-225.

Jensen, A. Verbal mediation and educational potential. *Psychology in the Schools*, 1966, **3**, 99-109.

Johnson, D. M., & Stratton, R. P. Evaluation of five methods of teaching concepts. *Journal of Educational Psychology*, 1966, **57**, 48-53.

Keppel, G., & Postman, L. Studies of learning to learn: III. Conditions of improvement in successive transfer tasks. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 260-267.

Laughlin, P. R., & Jordan, R. Selection strategies in conjunctive, disjunctive, and biconditional concept attainment. *Journal of Experimental Psychology*, 1967, **75**, 188-193.

McCandless, B. R. *Children and adolescents behavior and development.* New York: Holt, Rinehart and Winston, 1961.

Osler, S., & Fivel, M. W. Concept attainment: I. The role of age and intelligence in concept attainment by induction. *Journal of Experimental Psychology*, 1961, **62**, 1-8.

Pella, M. O. Concept learning. *Science Teacher*, 1966, **33**, 31-38.

Sechrest, L., & Wallace, J. Assimilation and utilization of information in concept attainment under varying conditions of information presentation. *Journal of Experimental Psychology*, 1962, **63**, 157-164.

Snow, C. E., & Rabinovitch, M. S. Disjunctive and conjunctive thinking in children. *Journal of Experimental Child Psychology*, 1969, **7**, 1-9.

Vaughan, G. M., & Corballis, M. C. Beyond tests of significance: Estimating strength of effects in selected ANOVA designs. *Psychological Bulletin*, 1969, **72**, 204-213.

(Received September 8, 1970)

Outstanding Educational Psychology Texts

ESSENTIALS OF LEARNING
Third Edition
By ROBERT M. W. TRAVERS, Western Michigan University
1972, approx. [illegible] pages, prob. [illegible]

MEASUREMENT AND EVALUATION IN TEACHING
Second Edition
By NORMAN E. GRONLUND, University of Illinois
1971, [illegible] pages, [illegible]

STATING BEHAVIORAL OBJECTIVES FOR CLASSROOM INSTRUCTION
By NORMAN E. GRONLUND, University of Illinois
1970, [illegible] pages, paper, [illegible]

Edited by MELVIN H. MARX, University of Missouri, Columbia

LEARNING: Processes
1969, [illegible] pages, [illegible]

LEARNING: Interactions
1970, [illegible] pages, [illegible]

LEARNING: Theories
1970, [illegible] pages, [illegible]

STUDIES IN ADOLESCENCE
A Book of Readings in Adolescent Development
Edited by ROBERT E. GRINDER, The University of Wisconsin
1969, [illegible] pages, paper, [illegible]

EDUCATIONAL PSYCHOLOGY
Third Edition
By GLENN MYERS BLAIR, R. STEWART JONES, and RAY H. SIMPSON, all, University of Illinois
1968, [illegible] pages, [illegible] [illegible] manual gratis

VOCATIONAL GUIDANCE AND CAREER DEVELOPMENT
Selected Readings
Second Edition
Edited by HERMAN J. PETERS, The Ohio State University, and JAMES C. HANSEN, State University of New York, Buffalo
1971, [illegible] pages, paper, [illegible]

For further information write to:
THE MACMILLAN COMPANY
Department C
Riverside, New Jersey 08075
In Canada, write to Collier-Macmillan Canada, Ltd., 1125B Leslie Street, Don Mills, Ontario, Canada

Journal of the Experimental Analysis of Behavior

The *Journal of the Experimental Analysis of Behavior* is primarily for the original publication of experiments relevant to the behavior of individual organisms. Review articles and theoretical papers will also be considered for publication. Invited book reviews appear on occasion.

EDITOR
Stanley S. Pliskoff, *University of Maine*

EXECUTIVE EDITOR
Victor G. Laties, *Univ. of Rochester*

ASSOCIATE EDITORS
Larry D. Byrd, *Harvard Medical School*
Lewis R. Gollub, *Univ. of Maryland*
Herbert M. Jenkins, *McMaster Univ.*
Howard Rachlin, *SUNY, Stony Brook*
Michael D. Zeiler, *Emory University*

REVIEW EDITOR
A. Charles Catania, *New York Univ.*

BOARD OF EDITORS
W. Baum, *Harvard University*
D. S. Blough, *Brown University*
J. J. Boren, *American University*
F. C. Clark, *Univ. of North Carolina*
M. R. D'Amato, *Rutgers University*
J. A. Dinsmoor, *Indiana University*
E. Fantino, *Univ. of Cal., La Jolla*
G. W. Farthing, *University of Maine*
R. M. Gilbert. *Alcoholism & Drug Addiction Res. Foundation*
D. F. Hake. *Anna State Hospital*
R. F. Hefferline, *Columbia University*
R. M. Herrick, *Air Dev. Ctr., Johnsville*
P. N. Hineline, *Temple University*
W. Hoods, *Walter Reed*
W. C. Holz, *Smith Kline & French*
D. Johnson, *Virginia Polytech. Inst.*
R. T. Kelleher, *Harvard Medical Sch.*
D. Kelly, *NYS Psychiatric Institute*
J. W. McKearney, *Worcester Foundation*
M. J. Marr, *Georgia Inst. of Tech.*
J. R. Millenson, *Univ. of Oxford*
D. E. Mintz, *CUNY, City College*
W. H. Morse, *Harvard Medical School*
A. Neuringer, *Reed College*
B. Ray, *Mass. General Hospital*
S. H. Revusky, *N. Illinois University*
G. S. Reynolds, *Univ. of Cal., La Jolla*
B. Schneider, *Columbia University*
J. Sherman, *University of Kansas*
C. Shimp, *University of Utah*
J. E. R. Staddon, *Duke University*
D. A. Stubbs, *University of Maine*
H. S. Terrace, *Columbia University*
T. Verhave, *CUNY, Queens College*
B. Weiss, *Univ. of Rochester*
D. R. Williams, *Univ. of Pennsylvania*
M. Yarczower, *Bryn Mawr College*
J. Zimmerman, *Indiana Univ. Med. School*

Of men and monkeys: the following were published in JEAB during 1971.

Birnbrauer, J. S. Effects of pairing stimuli with reinforcement on multiple schedule performance of children.

D'Amato, M. R. and O'Neill, W. Effect of delay-interval illumination on matching behavior in the capuchin monkey.

Dardano. J. F. Control of concurrent avoidance and appetitive behaviors by an indicator of shock proximity.

de Lorge, J. O. and Clark, F. C. Observing behavior in squirrel monkeys under a multiple schedule of reinforcement.

Findley, J. D., Robinson, W. W. and Gilliam, W. A restraint system for chronic study of the baboon.

Greene, W. A. and Sutor, L. T. Stimulus control of skin resistance responses on an escape-avoidance schedule.

Harrison, J. M., Downey, P., Segal, M. and Howe, M. Control of responding by location of auditory stimuli: rapid acquisition in monkey and rat.

Hutchinson, R. R., Renfrew, J. W. and Young, G. A. Effects of long-term shock and associated stimuli on aggressive and manual responses.

Krasnegor, N. A., Brady, J. V., and Findley, J. D. Second-order optional avoidance as a function of fixed-ratio requirements.

Mackay, H. A. and Brown, S. M. Teaching serial position sequences to monkeys with a delayed matching-to-sample procedure.

Miczek, K. A. and Grossman, S. P. Positive conditioned suppression: effects of CS duration.

Moody, D. B., Stebbins, W. C. and Iglauer, C. Auditory generalization gradients for response latency in the monkey.

Schmitt, D. R. and Marwell, G. Avoidance of risk as a determinant of cooperation.

Schmitt, D. R. and Marwell, G. Taking and the disruption of cooperation.

Stoddard, L. T. and Sidman, M. The removal and restoration of stimulus control.

Touchette, P. E. Transfer of stimulus control: measuring the moment of transfer.

Wilton, R. N. and Clements, R. O. The role of information in the emission of observing responses: a test of two hyoptheses.

Address subscription orders to Kay Dinsmoor, JEAB

Department of Psychology
Indiana University
Bloomington, Indiana 47401

Students $4 (payment with order)—Individuals (personal use) $8
Institutions $16